

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 16:48:07 ; Search time 7965.97 Seconds  
(without alignments)  
10902.766 Million cell updates/sec

Title: US-10-024-370-1  
Perfect score: 2123  
Sequence: 1 ctccgagcgsgagctcggtgatc.....gacctcatgcctcggaattc 2123

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_on:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_hgo\_hum:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2123	100.0	2123	1	CGI17592	Y17592 Corynebacte
2	2123	100.0	2123	6	AR202319	AR202319 Sequence
3	2123	100.0	2123	6	BD004741	BD004741 Replicabl
4	2111	99.4	332050	1	AP005276	AP005276 Corynebac
5	2111	99.4	349980	6	AX127145	AX127145 Sequence
6	1575	74.2	1575	6	AX066443	AX066443 Sequence
7	1473	69.4	1473	6	AR202320	AR202320 Sequence
8	1473	69.4	1473	6	AX121013	AX121013 Sequence
9	1473	69.4	1473	6	BD004742	BD004742 Replicabl
10	1473	69.4	1473	6	BD163130	BD163130 Novel pol
11	691.4	32.6	300750	1	AP005217	AP005217 Corynebac
12	515	24.3	3013	1	CGGLTNG	X66112 C.glutamicu
13	312.6	14.7	17910	1	AE006979	AE006979 Mycobacte
14	312.6	14.7	37630	1	MTCY31	Z73101 Mycobacteri
15	312.6	14.7	327650	1	BX248337	BX248337 Mycobacte
16	164.4	7.7	342300	1	MLEPRTN8	AE009634 Mycobacte
17	158.4	7.5	10029	1	AE009634	AE009634 Brucella
18	156.8	7.4	10029	1	AE014498	AE014498 Brucella
19	148.6	7.0	260050	1	SME591782	AL591782 Sinorhizo
20	142	6.7	36063	1	MLU15184	U15184 Mycobacteri
21	140.4	6.6	332635	1	AP003005	AP003005 Mesorhizo
22	136.6	6.4	298900	1	AP005937	AP005937 Bradyrhiz
23	132.6	6.2	294250	1	AP001517	AP001517 Bacillus
24	131	6.2	354	6	AX123526	AX123526 Sequence
25	131	6.2	354	6	BD165643	BD165643 Novel pol
26	128.4	6.0	7697	1	AE007944	AE007944 Agrobacte
27	128.4	6.0	9508	1	AE008976	AE008976 Agrobacte
28	124.8	5.9	6908	1	SPU59237	U59237 Synecococc
29	122.8	5.8	4776	1	AB031231	AB031231 Pseudomon
30	122.8	5.8	11976	1	AE004735	AE004735 Pseudomon
31	122.2	5.8	14401	1	AE001518	AE001518 Helicobac
32	121	5.7	304454	1	AE016956	AE016956 Enterococ
33	120.4	5.7	10740	1	AE012365	AE012365 Xanthomon
34	120.2	5.7	311321	1	NMA322491	AL162754 Neisseria
35	119.2	5.6	1053	6	BD092453	BD092453 Identific
36	119.2	5.6	14160	1	AE000604	AE000604 Helicobac
37	119	5.6	3777	1	AF139661	AF139661 Azospiril
38	118.8	5.6	300975	1	AE017013	AE017013 Bacillus
39	118.8	5.6	301200	1	AP005374	AP005374 Thermosyn
40	118.6	5.6	10377	1	AE002422	AE002422 Neisseria
41	118.6	5.6	349980	6	AX044030	AX044030 Sequence
42	117.2	5.5	291804	1	AE017039	AE017039 Bacillus
43	116.6	5.5	13369	1	AE006013	AE006013 Caulobact
44	113.4	5.3	300363	1	AE016781	AE016781 Pseudomon
45	110.8	5.2	14704	1	AE001970	AE001970 Deinococc

## ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE
CGI17592	CGI17592	Corynebacterium glutamicum accDA gene.	Y17592	Y17592.1	GI:14572581	accDA gene; acetyl-CoA carboxylase; carboxyltransferase.	Corynebacterium glutamicum	Kim,K., Peters-Wendisch,P.G., Sahm,H. and Eikmanns,B.J. unpublished
								1 (bases 1 to 2123)

AUTHORS Peters-Wendisch, P.G.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUN-1998) P.G. Peters-Wendisch, University of California at Berkeley, Dept. Plant and Microbial Biology, Kuecu Lab, 111 Koshland Hall, Berkeley CA 94720, USA  
COMMENT Related sequence x66112.  
FEATURES  
source  
1. 2123  
/organism="Corynebacterium glutamicum"  
/mol\_type="genomic DNA"  
/strain="ATCC 13032"  
/db\_xref="taxon:1718"  
508..1983  
/gene="accDA"  
508..1983  
/gene="accDA"  
/codon\_start=1  
/transl\_table=1  
/product="putative carboxyltransferase subunit of acetyl-CoA carboxylase"  
/protein\_id="CAC42827.1"  
/db\_xref="GI:14572582"  
/translation="MEKRFPTMVGMEHTSLALTLIDSLDPSHSWNETPYQDNLNGYAEETLRASAKCDSVLTGEVTEGIPIAVILSDPSFLGSLGTVASVRIKAIHRAETKLPILVSPASGGARMOEDNRAFVMWVSTAAVORHEALPIVAVIRNPTMGSLAMSGSSGHLTPAEPGAOTGLGPRVLETGHALPDGQANLIVKTVGIDIVSLP LQLRVAVAKTIVIPVEBATRSPPTPGVALPVEBALRSRDPQRIGIMETLGLADVRLSGARAGALSPAVRVALRIGRVRVILIGDRRTLSPOELRPRRISLARELNLPIVSIIDTSGAELSQAAEELGIASSIARTLSKIDAPLPTVSVIIGGGGGALMLPADLVVAENVAEMLSALPREGASAILFEDTHAAEILIERQCVQAHALLSQGLIDGIVAEHFEVEIIGTISNALSLELDNPERARDRSRFRFERLAAQ"  
BASE COUNT 460 a 574 c 592 g 497 t  
ORIGIN  
Query Match 100.0%; Score 2123; DB 1; Length 2123;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CTCGAGCGGGAGTGGGTATGGCGCACTCTTAGCAATGCCGGTTTAAATAAGCA 60  
DB 1 CTCGAGCGGGAGTGGGTATGGCGCACTCTTAGCAATGCCGGTTTAAATAAGCA 60  
QY 61 CTTATATGTTCTCCACCATCTGGCCGACCACTCTTAAGCAATGCCGGTTTAAATAAGCA 120  
DB 61 CTTATATGTTCTCCACCATCTGGCCGACCACTCTTAAGCAATGCCGGTTTAAATAAGCA 120  
QY 121 AACGTGTAATGTGAAGTACCTAATCAATGCAATGCGATGAGATTGGAAAATCTC 180  
DB 121 AACGTGTAATGTGAAGTACCTAATCAATGCAATGCGATGAGATTGGAAAATCTC 180  
QY 181 ACTGCCCCCAATATTCTTAAGTTAACTTAAAGTAGTGTCTTACTGATTTATTAAGT 240  
DB 181 ACTGCCCCCAATATTCTTAAGTTAACTTAAAGTAGTGTCTTACTGATTTATTAAGT 240  
QY 241 TCCGATCTACCCCTCTTACCCTCGGAATACCCCTTTTGCAGAAAGATTGCAACACACA 300  
DB 241 TCCGATCTACCCCTCTTACCCTCGGAATACCCCTTTTGCAGAAAGATTGCAACACACA 300  
QY 241 TCCGATCTACCCCTCTTACCCTCGGAATACCCCTTTTGCAGAAAGATTGCAACACACA 300  
DB 241 TCCGATCTACCCCTCTTACCCTCGGAATACCCCTTTTGCAGAAAGATTGCAACACACA 300  
QY 301 GTGCAATAGTTAAAGGGCTTCAACAGTCAATTCGTCCGGTTTAAAGCTATGTTCCGG 360  
DB 301 GTGCAATAGTTAAAGGGCTTCAACAGTCAATTCGTCCGGTTTAAAGCTATGTTCCGG 360  
QY 301 GTGCAATAGTTAAAGGGCTTCAACAGTCAATTCGTCCGGTTTAAAGCTATGTTCCGG 360  
DB 301 GTGCAATAGTTAAAGGGCTTCAACAGTCAATTCGTCCGGTTTAAAGCTATGTTCCGG 360  
QY 361 ACGTCTAGGCAAAAAGTAGTTTGTAGATGAAGCAATATCCGTCAATTTTTCGCA 420  
DB 361 ACGTCTAGGCAAAAAGTAGTTTGTAGATGAAGCAATATCCGTCAATTTTTCGCA 420  
QY 421 TCGATAGCTAAATTTGGGCTTAGATCTCCGCTTAATAGGTATGAGAGACATTGGA 480  
DB 421 TCGATAGCTAAATTTGGGCTTAGATCTCCGCTTAATAGGTATGAGAGACATTGGA 480  
QY 421 TCGATAGCTAAATTTGGGCTTAGATCTCCGCTTAATAGGTATGAGAGACATTGGA 480  
DB 421 TCGATAGCTAAATTTGGGCTTAGATCTCCGCTTAATAGGTATGAGAGACATTGGA 480  
QY 481 ATTATTTAAACAAGCCATTTTTCGGCGTGAAGAGCTTTTCCGACTATGGTGTGGGC 540  
DB 481 ATTATTTAAACAAGCCATTTTTCGGCGTGAAGAGCTTTTCCGACTATGGTGTGGGC 540  
481 ATTATTTAAACAAGCCATTTTTCGGCGTGAAGAGCTTTTCCGACTATGGTGTGGGC 540

QY 541 ATGGAACACACTTCAGATTGAGCGCTCATPAGACTGGTTTGGACCCCTGACAGCTTCATT 600  
DB 541 ATGGAACACACTTCAGATTGAGCGCTCATPAGACTGGTTTGGACCCCTGACAGCTTCATT 600  
QY 601 TCTTGAATGAACTCCCAATATGACCAACTCAATCAAGCTATGACAGACTTTGGAG 660  
DB 601 TCTTGAATGAACTCCCAATATGACCAACTCAATCAAGCTATGACAGACTTTGGAG 660  
QY 661 CGGGCTCGAAGCAAGGCAAAATCGATGAATCGGTAAATTTCTGAGAGAGGACCCGTGAG 720  
DB 661 CGGGCTCGAAGCAAGGCAAAATCGATGAATCGGTAAATTTCTGAGAGAGGACCCGTGAG 720  
QY 721 GGCATTCGGTAGCCGTTATTTTTCGATATTTTCTCTCTCGGCGGTTCTTTGGGACG 780  
DB 721 GGCATTCGGTAGCCGTTATTTTTCGATATTTTCTCTCTCGGCGGTTCTTTGGGACG 780  
QY 781 GTGCGGTGCGTGGCATCATGAAGGCGATTCAACCGGCCACAGAGCTGAATCTCCACTG 840  
DB 781 GTGCGGTGCGTGGCATCATGAAGGCGATTCAACCGGCCACAGAGCTGAATCTCCACTG 840  
QY 841 CTGGTCTCCCTGCTTCCGGTGTGCGCGCATGCAAGAGACAAATGAGCTTTTGTATG 900  
DB 841 CTGGTCTCCCTGCTTCCGGTGTGCGCGCATGCAAGAGACAAATGAGCTTTTGTATG 900  
QY 901 ATGGTGTCAATACCGCGGCTGTGACGCGTCAACGCGAGGCGGATTTGCGGTTCTG 960  
DB 901 ATGGTGTCAATACCGCGGCTGTGACGCGTCAACGCGAGGCGGATTTGCGGTTCTG 960  
QY 961 TATTGCGCAATCCACGATGAGTGTGCGCGCATGACCTGTGAGGATTCATGCGGCATCTC 1020  
DB 961 TATTGCGCAATCCACGATGAGTGTGCGCGCATGACCTGTGAGGATTCATGCGGCATCTC 1020  
QY 1021 ACTTTTCCGGAACCCGCGCGCGAGATAGTTTCTGAGTCTCTGCGGTGTGAGATTACC 1080  
DB 1021 ACTTTTCCGGAACCCGCGCGCGAGATAGTTTCTGAGTCTCTGCGGTGTGAGATTACC 1080  
QY 1081 ACTGGGCGATCGCTTCAGACGCGTGTGACAGGCGGAGAAATTTGTGAAAACCTGTGTG 1140  
DB 1081 ACTGGGCGATCGCTTCAGACGCGTGTGACAGGCGGAGAAATTTGTGAAAACCTGTGTG 1140  
QY 1141 ATTGATGAAATTTGTGTGCGCACTCAATTCGTGACAGCGGTGTGCAAAACCTCAAGTT 1200  
DB 1141 ATTGATGAAATTTGTGTGCGCACTCAATTCGTGACAGCGGTGTGCAAAACCTCAAGTT 1200  
QY 1201 ATTGAGCCGGTAGAGGCAAGGATCGTTTCTTCAACACTCTGCGCGTGTGCACTTCCG 1260  
DB 1201 ATTGAGCCGGTAGAGGCAAGGATCGTTTCTTCAACACTCTGCGCGTGTGCACTTCCG 1260  
QY 1261 GTGATGAGGCGAATTTGCGGCTTCGTGACCCGAGAGGCGTGGAAATCGGGAGATTATG 1320  
DB 1261 GTGATGAGGCGAATTTGCGGCTTCGTGACCCGAGAGGCGTGGAAATCGGGAGATTATG 1320  
QY 1321 GAAACGTTGGGGGAGAGCGTGTCAAGCTTCTGTGTCGCGTGTGCGCATTTAGGCCG 1380  
DB 1321 GAAACGTTGGGGGAGAGCGTGTCAAGCTTCTGTGTCGCGTGTGCGCATTTAGGCCG 1380  
QY 1381 GCTGTGCGCTTGTGCGCTGTGCGCGCATGCGGGGCGGCGCGTGTGTGATTTGGGCAAGAT 1440  
DB 1381 GCTGTGCGCTTGTGCGCTGTGCGCGCATGCGGGGCGGCGCGTGTGTGATTTGGGCAAGAT 1440  
QY 1441 CGCGGCTTCAACGCTTGGGCGGAGAGCTGCTTTTTCGCGTGTGCGCATTTGCTGGCG 1500  
DB 1441 CGCGGCTTCAACGCTTGGGCGGAGAGCTGCTTTTTCGCGTGTGCGCATTTGCTGGCG 1500  
QY 1501 CGCAGCTAAACCTGCGGATGTGTCCATCATGACACTTCGCGCGCGGATTTGTCCAG 1560  
DB 1501 CGCAGCTAAACCTGCGGATGTGTCCATCATGACACTTCGCGCGCGGATTTGTCCAG 1560  
QY 1561 GCGGCTGAGAGCTGTGCGATGTGCAATGCGAATTTGCGGCACTTTGTCCAAAGCTTATGAC 1620  
DB 1561 GCGGCTGAGAGCTGTGCGATGTGCAATGCGAATTTGCGGCACTTTGTCCAAAGCTTATGAC 1620



QY 1621 GCTCCCTCCCAACCGCTTTCGATTAATGTCAGGCGCTTGGCGGCTGCGC 1680  
 Db 1621 GCTCCCTCCCAACCGCTTTCGATTAATGTCAGGCGCTTGGCGGCTGCGC 1680  
 QY 1681 ATGCTGCTCCCGCATCTGCTGTCACGCGCGGAAAAGCGCTGCTGCTCCGATTCGACCA 1740  
 Db 1681 ATGCTGCTCCCGCATCTGCTGTCACGCGCGGAAAAGCGCTGCTGCTCCGATTCGACCA 1740  
 QY 1741 GAGGCGCTCGGCGCATCTCTTCCGCGACACCAACGCGGGAATCATAGAGCG 1800  
 Db 1741 GAGGCGCTCGGCGCATCTCTTCCGCGACACCAACGCGGGAATCATAGAGCG 1800  
 QY 1801 CAAGCGTGCAGCGCGACCACTTTTAAGCGCAAGGCTTATCGACGCGATCGCGCA 1860  
 Db 1801 CAAGCGTGCAGCGCGACCACTTTTAAGCGCAAGGCTTATCGACGCGATCGCGCA 1860  
 QY 1861 ACCGACATCTTGTGTAAGAAATTCGCGACATGACAGCGCTTCCGATTTGAA 1920  
 Db 1861 ACCGACATCTTGTGTAAGAAATTCGCGACATGACAGCGCTTCCGATTTGAA 1920  
 QY 1921 AACCATCGGAGAGGCGGAGCGGACAGTCCCTTACACAGATTGAGGCTTAAGCGCA 1980  
 Db 1921 AACCATCGGAGAGGCGGAGCGGACAGTCCCTTACACAGATTGAGGCTTAAGCGCA 1980  
 QY 1981 TAAAGAAATTAATGCTGCTGATCAATCGATGATGAACACAGGCTACGCGCAGACGTG 2040  
 Db 1981 TAAAGAAATTAATGCTGCTGATCAATCGATGATGAACACAGGCTACGCGCAGACGTG 2040  
 QY 2041 GTGGCGGAAACCTTCAGGCGCGTGAAGCACTCTGCGGAAATGTCAGTGAAGCGTTC 2100  
 Db 2041 GTGGCGGAAACCTTCAGGCGCGTGAAGCACTCTGCGGAAATGTCAGTGAAGCGTTC 2100  
 QY 2101 GCCGACCTTCATGCGCTGGAATTC 2123  
 Db 2101 GCCGACCTTCATGCGCTGGAATTC 2123

RESULT 2  
 LOCUS AR202319 2123 bp DNA linear PAT 20-APR-2002  
 DEFINITION Sequence 1 from patent US 6361986.  
 ACCESSION AR202319  
 VERSION AR202319.1 GI:20256858  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 2123)  
 AUTHORS Tili, Y., Birkmann, B., Eggeling, L., Sahm, H. and Mockel, B.  
 TITLE Process for the preparation of L-amino acids by fermentation and  
 nucleotide sequences coding for the accD gene  
 JOURNAL Patent: US 6361986-A 1 26-MAR-2002;  
 FEATURES  
 source 1..2123  
 location/Qualifiers  
 BASE COUNT 460 a 574 c 592 g 497 t

Query Match 100.0%; Score 2123; DB 6; Length 2123;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGAGCGGAGTGGTGTGATCGGCGCATCTCTAAGCAATGCGGCTTTAAATAAGCA 60  
 Db 1 CTCGAGCGGAGTGGTGTGATCGGCGCATCTCTAAGCAATGCGGCTTTAAATAAGCA 60  
 QY 61 CTATATGTTTCTACCAATCTGGCGACGACGACGAGTATGTTGCGATCAAGCTTA 120  
 Db 61 CTATATGTTTCTACCAATCTGGCGACGACGACGAGTATGTTGCGATCAAGCTTA 120  
 QY 121 AACGTGTAATGTAGTAACTTAATCTCAATGCAATGCGATGAGTGGATTTGAAAATCTC 180  
 Db 121 AACGTGTAATGTAGTAACTTAATCTCAATGCAATGCGATGAGTGGATTTGAAAATCTC 180

QY 181 ACTCCCCCAATATCTTAATCTTAACTTAAAGTATGTTTTACTGCAATTTATAAAGT 240  
 Db 181 ACTCCCCCAATATCTTAATCTTAACTTAAAGTATGTTTTACTGCAATTTATAAAGT 240  
 QY 241 TCCGATCTAACCCCTCTTTTACCAGCAATACCCCTTTTGGCAAGATGCAACACACA 300  
 Db 241 TCCGATCTAACCCCTCTTTTACCAGCAATACCCCTTTTGGCAAGATGCAACACACA 300  
 QY 301 GTGCAATAGTTAACGGGCTTCAACAGTCAATCTGTCGGGTTTAAAGCTATGTTCCG 360  
 Db 301 GTGCAATAGTTAACGGGCTTCAACAGTCAATCTGTCGGGTTTAAAGCTATGTTCCG 360  
 QY 361 GTGCAATAGTTAACGGGCTTCAACAGTCAATCTGTCGGGTTTAAAGCTATGTTCCG 420  
 Db 361 GTGCAATAGTTAACGGGCTTCAACAGTCAATCTGTCGGGTTTAAAGCTATGTTCCG 420  
 QY 421 TCGATAGCTTAATTTGAGGCTTATGATCTTCGCTCTAATATAGTATGACAGATTCGA 480  
 Db 421 TCGATAGCTTAATTTGAGGCTTATGATCTTCGCTCTAATATAGTATGACAGATTCGA 480  
 QY 481 ATTAATTAACAAAGCCATTTTTCGCGCGTGAAGACGTTTTCCGATATGATGCGGC 540  
 Db 481 ATTAATTAACAAAGCCATTTTTCGCGCGTGAAGACGTTTTCCGATATGATGCGGC 540  
 QY 541 ATGGAACACACTTACGATTCAGCTGATGACCTGCTGATGCTGCTTTTGGACCTGACCTTCAT 600  
 Db 541 ATGGAACACACTTACGATTCAGCTGATGACCTGCTGATGCTGCTTTTGGACCTGACCTTCAT 600  
 QY 601 TCTTGAATGAACCTCCCAATATGACAACTCAATCAAGGCTATGACAGACCTTGGAG 660  
 Db 601 TCTTGAATGAACCTCCCAATATGACAACTCAATCAAGGCTATGACAGACCTTGGAG 660  
 QY 661 CGGCTCGAAGCAAGGCGCAATGCGATGATCGTAAATCTGTAAGAAAGGCAACCGTGGAG 720  
 Db 661 CGGCTCGAAGCAAGGCGCAATGCGATGATCGTAAATCTGTAAGAAAGGCAACCGTGGAG 720  
 QY 721 GGCATTCGCGTATGCGCTTATTTTTCGATTTTCTTCTGCGCGCTTCTTTGGGACG 780  
 Db 721 GGCATTCGCGTATGCGCTTATTTTTCGATTTTCTTCTGCGCGCTTCTTTGGGACG 780  
 QY 781 GTGCGTGTGCGCATGATGAAAGGCAATTCACGCGCGACAGACGTGAATCTCCACAG 840  
 Db 781 GTGCGTGTGCGCATGATGAAAGGCAATTCACGCGCGACAGACGTGAATCTCCACAG 840  
 QY 841 CTGATCTCCCTGCTTCCGATGCTGCGGCGCATGACAGAAACAATGACCTTTTGTCAAG 900  
 Db 841 CTGATCTCCCTGCTTCCGATGCTGCGGCGCATGACAGAAACAATGACCTTTTGTCAAG 900  
 QY 901 ATGCTGTCAATPACCGCGCTGTGTGACAGCTGACCGCGAGGCGCATTTCCGTCGTG 960  
 Db 901 ATGCTGTCAATPACCGCGCTGTGTGACAGCTGACCGCGAGGCGCATTTCCGTCGTG 960  
 QY 961 TATTGGCGCAATCCACAGATGGTGGCGCAATGCGCTGTGGGTTCAATCGGCAATCTC 1020  
 Db 961 TATTGGCGCAATCCACAGATGGTGGCGCAATGCGCTGTGGGTTCAATCGGCAATCTC 1020  
 QY 1021 ACTTTTGGGAACCGCGCGCAGATAGTTCCTGAGTCTCGCTGCTGCTGAGTAAAC 1080  
 Db 1021 ACTTTTGGGAACCGCGCGCAGATAGTTCCTGAGTCTCGCTGCTGCTGAGTAAAC 1080  
 QY 1081 ACTGGCGATGCGCTTCCAGACGCTGTGTGACAGAGCGGAGAAATTTGTGAAAACTGTGTG 1140  
 Db 1081 ACTGGCGATGCGCTTCCAGACGCTGTGTGACAGAGCGGAGAAATTTGTGAAAACTGTGTG 1140  
 QY 1141 ATTGAATGAATTTGTGTGCGCATCTCAATGCGTCAAGGCTTAAAGCTTAAAGTT 1200  
 Db 1141 ATTGAATGAATTTGTGTGCGCATCTCAATGCGTCAAGGCTTAAAGCTTAAAGTT 1200  
 QY 1201 ATTGAATGAATTTGTGTGCGCATCTCAATGCGTCAAGGCTTAAAGCTTAAAGTT 1260  
 Db 1201 ATTGAATGAATTTGTGTGCGCATCTCAATGCGTCAAGGCTTAAAGCTTAAAGTT 1260

QY 1261 GTATGAGAGGAGATTGCGCTTCTGTAACCCGAGAGGCTGTGAATCGGGAGATTATG 1320  
 Db 1261 GTATGAGAGGAGATTGCGCTTCTGTAACCCGAGAGGCTGTGAATCGGGAGATTATG 1320  
 QY 1321 GAAACGTTGGGGGAGAGAGTGTCAAGCTTTCTGAGCGCTGTGCGCATTTAGACCCG 1380  
 Db 1321 GAAACGTTGGGGGAGAGAGTGTCAAGCTTTCTGAGCGCTGTGCGCATTTAGACCCG 1380  
 QY 1381 GCTGTGGCGCTTGTGCGCGCGCATCGGGGGCGCGCGCGGTGTGATTTGGGACAGAT 1440  
 Db 1381 GCTGTGGCGCTTGTGCGCGCGCATCGGGGGCGCGCGCGGTGTGATTTGGGACAGAT 1440  
 QY 1441 CGCGCTTCAAGCTTGTGGCGCGAGAGCTGGCTTTTGGCGCTGTGGCATTTCCCTGGCG 1500  
 Db 1441 CGCGCTTCAAGCTTGTGGCGCGAGAGCTGGCTTTTGGCGCTGTGGCATTTCCCTGGCG 1500  
 QY 1501 CGCGAGCTAAACCTGCGCATGTGTCAATCAATCAACCTCGCGCGCATTTGTGCGAG 1560  
 Db 1501 CGCGAGCTAAACCTGCGCATGTGTCAATCAATCAACCTCGCGCGCATTTGTGCGAG 1560  
 QY 1561 GCGGCTGAGAGGCTCGGATCGCAAGCTGATTTGGCGCGCTGTGCGCATTTGCGAG 1620  
 Db 1561 GCGGCTGAGAGGCTCGGATCGCAAGCTGATTTGGCGCGCTGTGCGCATTTGCGAG 1620  
 QY 1621 GCTCCCTCCCACTGCTTGTGCTATTTGTGATTTGTGAGGCGCTGTGCGCATTTGCGAG 1680  
 Db 1621 GCTCCCTCCCACTGCTTGTGCTATTTGTGATTTGTGAGGCGCTGTGCGCATTTGCGAG 1680  
 QY 1681 ATGTGCTCCCGCATTTGTGCTATTTGTGATTTGTGAGGCGCTGTGCGCATTTGCGAG 1740  
 Db 1681 ATGTGCTCCCGCATTTGTGCTATTTGTGATTTGTGAGGCGCTGTGCGCATTTGCGAG 1740  
 QY 1741 GAGGCGCTGCGCATTTGTGCTATTTGTGATTTGTGAGGCGCTGTGCGCATTTGCGAG 1800  
 Db 1741 GAGGCGCTGCGCATTTGTGCTATTTGTGATTTGTGAGGCGCTGTGCGCATTTGCGAG 1800  
 QY 1801 CAAGGCGCTGCGCATTTGTGCTATTTGTGATTTGTGAGGCGCTGTGCGCATTTGCGAG 1860  
 Db 1801 CAAGGCGCTGCGCATTTGTGCTATTTGTGATTTGTGAGGCGCTGTGCGCATTTGCGAG 1860  
 QY 1861 ACCGAGCACTTGTGTAAGAAATTTCTCGGCAACAATCAACCACTGCGAAATTTGAT 1920  
 Db 1861 ACCGAGCACTTGTGTAAGAAATTTCTCGGCAACAATCAACCACTGCGAAATTTGAT 1920  
 QY 1921 AACGATCGGAGAGGCGCGAGCGAGCGAGCTTCAACAGATTTGAGCGGAG 1980  
 Db 1921 AACGATCGGAGAGGCGCGAGCGAGCGAGCTTCAACAGATTTGAGCGGAG 1980  
 QY 1981 TAAAGAAAATTTATGCGCTGATCAAAATCGATGATGAACACAGGGTACGGACAGAGTGG 2040  
 Db 1981 TAAAGAAAATTTATGCGCTGATCAAAATCGATGATGAACACAGGGTACGGACAGAGTGG 2040  
 QY 2041 GTGGCGGAGAACCTTCAAGGCGGTAAAGCAAGCTTGTGCGGAAATGCTCACTGACAGAGTGC 2100  
 Db 2041 GTGGCGGAGAACCTTCAAGGCGGTAAAGCAAGCTTGTGCGGAAATGCTCACTGACAGAGTGC 2100  
 QY 2101 GCCGACCTTCAATGCTGTGAATTC 2123  
 Db 2101 GCCGACCTTCAATGCTGTGAATTC 2123

RESULT 3  
 BD004741 2123 bp DNA linear PAT 31-JAN-2002  
 LOCUS  
 DEFINITION Replicable DNA, amino acid sequence, *Corynebacterium* microorganism,  
 shuttle vector, and process for producing L-amino acid.  
 ACCESSION BD004741  
 VERSION BD004741.1 GI:18632702  
 KEYWORDS UP 2001008693-A/1.  
 SOURCE  
 ORGANISM  
*Corynebacterineae*; *Corynebacteriaceae*; *Corynebacterium*.

REFERENCE 1 (bases 1 to 2123)  
 AUTHORS Tilke, I., Eggert, L., Sickmans, B., Zamu, H. and Meck, V.  
 TITLE Replicable DNA, amino acid sequence, *Corynebacterium* microorganism,  
 shuttle vector, and process for producing L-amino acid  
 JOURNAL Patent: JP 2001008693-A 1 16-JAN-2001;  
 DEUSSA HUELS AG, FORSCHUNGSZENTRUM JUELICH GMBH  
 OS *Corynebacterium* glutamicum  
 PN JP 2001008693-A/1  
 PD 16-JAN-2001  
 PR 24-MAY-2000 JP 200053547  
 PR 27-MAY-1999 DE 19924365.4  
 PI IVENNE TILKE, LOTHAR EGGERT, BERNHARD EICKMANS, HERMANN ZAMU,  
 PI VETTYNA MECKEL  
 PC C12N15/09, C12N1/21, C12P13/04, C12P13/06, C12P13/08, C12P13/12, C12P13/20, C12P21/02, C12N15/09, C12R1/15, C12N1/21, C12R1/15, C12P13/06, C12R1/15, C12P13/08, C12R1/15, C12P13/12, C12R1/15, C12P13/20, C12R1/15, C12P21/02, C12R1/15, C12N15/00, C12N15/00, C12R1/15)  
 PC (C12P13/04, C12R1/15), (C12P13/06, C12R1/15), (C12P13/08, C12R1/15), PC (C12P13/12, C12R1/15), (C12P13/20, C12R1/15), (C12P21/02, C12R1/15), PC C12N15/00, C12N15/00, C12R1/15)  
 CC (C12N15/00, C12R1/15)  
 FT gene  
 FEATURES  
 source  
 1. 2123  
 /organism="Corynebacterium glutamicum"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:1718"  
 BASE COUNT 460 a 574 c 592 g 497 t  
 ORIGIN  
 Query Match 100.0%; Score 2123; DB 6; Length 2123;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CTGAGGCGGAGAGTGGATGCGGCACTCTTAAGCAATCGGCTTTAAATTAAGCAA 60  
 Db 1 CTGAGGCGGAGAGTGGATGCGGCACTCTTAAGCAATCGGCTTTAAATTAAGCAA 60  
 QY 61 CTATATGTTTCTCACCACATCTGGCGGAGCAGCAGCAAGATATGTTGATCAACACTA 120  
 Db 61 CTATATGTTTCTCACCACATCTGGCGGAGCAGCAGCAAGATATGTTGATCAACACTA 120  
 QY 121 AACGTGATGTAAGTATCTTAACCTCACTGCAATTCGATATGCGATTTGGAAAACTC 180  
 Db 121 AACGTGATGTAAGTATCTTAACCTCACTGCAATTCGATATGCGATTTGGAAAACTC 180  
 QY 181 ACTGCGGCAATTAATCTTAACCTTAATTAAGTATGTTTACCTGATTTAAAGT 240  
 Db 181 ACTGCGGCAATTAATCTTAACCTTAATTAAGTATGTTTACCTGATTTAAAGT 240  
 QY 241 TCCGATCTACCCCTCTTTACCCGAAATACCCCTTTTCAAGATTTGCAAAACACACA 300  
 Db 241 TCCGATCTACCCCTCTTTACCCGAAATACCCCTTTTCAAGATTTGCAAAACACACA 300  
 QY 301 GTGCAATGATTAACGGGCTTCAACGTCACCATTTCTGCGGTTTGGCTATGTTGGG 360  
 Db 301 GTGCAATGATTAACGGGCTTCAACGTCACCATTTCTGCGGTTTGGCTATGTTGGG 360  
 QY 361 ACGTCTAGGCAAAAAGTATTTGTGAGTGAAGAACGATATTCGTAATTTTACGCA 420  
 Db 361 ACGTCTAGGCAAAAAGTATTTGTGAGTGAAGAACGATATTCGTAATTTTACGCA 420  
 QY 421 TCGATACCTTAATTTGGGCTTGAATCTTCCGCTCTTAATAGATGATCAAGAGATTCGA 480  
 Db 421 TCGATACCTTAATTTGGGCTTGAATCTTCCGCTCTTAATAGATGATCAAGAGATTCGA 480  
 QY 481 ATTAATTAACAAAGCAATTTTGGCGGTGAGAGAGGCTTTTCCGACTATGTTGGGCG 540  
 Db 481 ATTAATTAACAAAGCAATTTTGGCGGTGAGAGAGGCTTTTCCGACTATGTTGGGCG 540

QY 541 ATGGAACACACTTACATTCATGACGCTCATAGACTCGGTTTGGACCCCTGACAGCTTCATT 600  
 DB 541 ATGGAACACACTTACATTCATGACGCTCATAGACTCGGTTTGGACCCCTGACAGCTTCATT 600  
 QY 601 TTTTGGAAATGAAATCTCCCAATATGCAACTTCATCAAGGCTATGACAGACCTTTGAG 660  
 DB 601 TTTTGGAAATGAAATCTCCCAATATGCAACTTCATCAAGGCTATGACAGACCTTTGAG 660  
 QY 661 CGGGCTCGAAGCAAGGCGCAATGCGATGTAATCGGTAACTAGAGAAAGGCAACCGTGGAG 720  
 DB 661 CGGGCTCGAAGCAAGGCGCAATGCGATGTAATCGGTAACTAGAGAAAGGCAACCGTGGAG 720  
 QY 721 GGCATTTCCGGTAGCGGTAATTTTGTCCGATTTTCTCTCGGCGGTTCTTTGGGACAG 780  
 DB 721 GGCATTTCCGGTAGCGGTAATTTTGTCCGATTTTCTCTCGGCGGTTCTTTGGGACAG 780  
 QY 781 GTCCGCTGGGTGGCATCATGAAAGGCGATTACCCGCGCAAGAGCTGAACTCCCATG 840  
 DB 781 GTCCGCTGGGTGGCATCATGAAAGGCGATTACCCGCGCAAGAGCTGAACTCCCATG 840  
 QY 841 CTGGTCTCCCTCGCTCCGGTGGTGGCGCATGACAGGAAGACATCGACTTTTGTCAAG 900  
 DB 841 CTGGTCTCCCTCGCTCCGGTGGTGGCGCATGACAGGAAGACATCGACTTTTGTCAAG 900  
 QY 901 ATGGTGTCCATTAACCGCGCTGTGACAGCTCACCGGAGCGCATTTTGGCTTCTGGTG 960  
 DB 901 ATGGTGTCCATTAACCGCGCTGTGACAGCTCACCGGAGGCCATTTTGGCTTCTGGTG 960  
 QY 961 TATTTGCGCAATCCCAAGATGAGGCGCGCATGAGCTCGTGGGAGTTTCACTGGGCAATCTC 1020  
 DB 961 TATTTGCGCAATCCCAAGATGAGGCGCGCATGAGCTCGTGGGAGTTTCACTGGGCAATCTC 1020  
 QY 1021 ACTTTTGGGGAACCCGGGCGGAGATAGGTTTCTCGGCTCTCGGCTGTGAGATTAAC 1080  
 DB 1021 ACTTTTGGGGAACCCGGGCGGAGATAGGTTTCTCGGCTCTCGGCTGTGAGATTAAC 1080  
 QY 1081 ACTGGGCATATGCGCTTCCAGACGCTGTGACAGCGGAGAAATTTGGTGAATCTGCTGTG 1140  
 DB 1081 ACTGGGCATATGCGCTTCCAGACGCTGTGACAGCGGAGAAATTTGGTGAATCTGCTGTG 1140  
 QY 1141 ATTGATGGAATTTGTCTGCGCACTTCCAAATGCGTGCAGCGGTGGCAAAAACCTCAAGGT 1200  
 DB 1141 ATTGATGGAATTTGTCTGCGCACTTCCAAATGCGTGCAGCGGTGGCAAAAACCTCAAGGT 1200  
 QY 1201 ATTGACCGCGTAGAGGCAACGATGTTTCTCCAACTCTGGGCTGACATCTTCCG 1260  
 DB 1201 ATTGACCGCGTAGAGGCAACGATGTTTCTCCAACTCTGGGCTGACATCTTCCG 1260  
 QY 1261 GTGATGAGAGGCAATGCGCTTCTGTGACCGGAGAGGCTGGAATCGGGAGATTATG 1320  
 DB 1261 GTGATGAGAGGCAATGCGCTTCTGTGACCGGAGAGGCTGGAATCGGGAGATTATG 1320  
 QY 1321 GAAACGTTGGGGGCGAGAGCTGCTCAAGCTTCTGTGTCGCGTGTGCGCATTTGAGCCCG 1380  
 DB 1321 GAAACGTTGGGGGCGAGAGCTGCTCAAGCTTCTGTGTCGCGTGTGCGCATTTGAGCCCG 1380  
 QY 1381 GCTGTGCGCGTTCCTCTGCGCGCATCGGGGGCCGCGCGGTGCTCTAATGGGAGAGAT 1440  
 DB 1381 GCTGTGCGCGTTCCTCTGCGCGCATCGGGGGCCGCGCGGTGCTCTAATGGGAGAGAT 1440  
 QY 1441 CGCCGCTTCAGCGTTCGGGCGCGAGAGCTGCGTTTGGCGCGTGTGCAATTCGTGGCG 1500  
 DB 1441 CGCCGCTTCAGCGTTCGGGCGCGAGAGCTGCGTTTGGCGCGTGTGCAATTCGTGGCG 1500  
 QY 1501 CGCGAGCTAAACCTGCGCATGCTTCATCATGACACCTCGCGGCGGAATGTGCGAG 1560  
 DB 1501 CGCGAGCTAAACCTGCGCATGCTTCATCATGACACCTCGCGGCGGAATGTGCGAG 1560  
 QY 1561 GCGGCTGAGAGCTGGGATGCGAAGCTCGATTTGGCGGCACTTGTCCAAAGCTTATTCAC 1620  
 DB 1561 GCGGCTGAGAGCTGGGATGCGAAGCTCGATTTGGCGGCACTTGTCCAAAGCTTATTCAC 1620  
 QY 1621 GCTCCCTCTCCACCGTTCGGTTCATTAATGCTCAGGAGGTTGGCGGTGGCGCTGGCC 1680

DB 1621 GCTCCCTCTCCACCGTTCGGTTCATTAATGCTCAGGAGGTTGGGCGCGCTGGCC 1680  
 QY 1681 ATGCTGCGCGCGATCTGCTTACCGCGGCGGAAAACGGTGGCTGTCCGATTTGCCACCA 1740  
 DB 1681 ATGCTGCGCGCGATCTGCTTACCGCGGCGGAAAACGGTGGCTGTCCGATTTGCCACCA 1740  
 QY 1741 GAGGGCGGCTGGGCGATCTCTTCCGCAACCAACCGCGCGGAAATCATAGAGCA 1800  
 DB 1741 GAGGGCGGCTGGGCGATCTCTTCCGCAACCAACCGCGCGGAAATCATAGAGCA 1800  
 QY 1801 CAAGCGTGCAGGCGCGACGCACTTTTAAGCCAAAGGCTTATGACGGGATCGTCCCGAA 1860  
 DB 1801 CAAGCGTGCAGGCGCGACGCACTTTTAAGCCAAAGGCTTATGACGGGATCGTCCCGAA 1860  
 QY 1861 ACCGAGCACTTTGTTGAAGAAATTTCTCGGCAACATCAAGCAAGCCCTCTCGGAATTGAT 1920  
 DB 1861 ACCGAGCACTTTGTTGAAGAAATTTCTCGGCAACATCAAGCAAGCCCTCTCGGAATTGAT 1920  
 QY 1921 AACATTCGGGAGAGGCGGGAACGCGACAGTGGCTTACAGATTTGAGCGTTAGCGAG 1980  
 DB 1921 AACATTCGGGAGAGGCGGGAACGCGACAGTGGCTTACAGATTTGAGCGTTAGCGAG 1980  
 QY 1981 TAAAGAAATTAATGCGCTGATCAATTCGATGTAACACAGGCTACGCGCAGACAGTGG 2040  
 DB 1981 TAAAGAAATTAATGCGCTGATCAATTCGATGTAACACAGGCTACGCGCAGACAGTGG 2040  
 QY 2041 GTGGCGGAACCTTCAGGCGCGTAAAGAGCTCTTGGGGAATGTCACGTGACGAGTCC 2100  
 DB 2041 GTGGCGGAACCTTCAGGCGCGTAAAGAGCTCTTGGGGAATGTCACGTGACGAGTCC 2100  
 QY 2101 GCCGACCTTCATGCTCGGAATTC 2123  
 DB 2101 GCCGACCTTCATGCTCGGAATTC 2123

RESULT 4  
 AP005276/c 332050 bp DNA linear BCT 08-AUG-2002  
 LOCUS Corynebacterium glutamicum ATCC 13032 DNA, complete genome, section 3/10.  
 DEFINITION AP005276 BA000036  
 ACCESSION AP005276.1 GI:21323419  
 VERSION  
 KEYWORDS  
 SOURCE Corynebacterium glutamicum ATCC 13032  
 ORGANISM Corynebacterium glutamicum ATCC 13032  
 Bacteria; Actinobacteriales; Actinomycetales; Corynebacteriaceae; Corynebacterium.  
 REFERENCE 1 Nakagawa, S.  
 Complete genomic sequence of Corynebacterium glutamicum ATCC 13032  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 332050)  
 AUTHORS Nakagawa, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (24-MAY-2002) Satoshi Nakagawa, Kyowa Hakko Kogyo Co. Ltd., Tokyo Research Laboratories; 3-6-6, Asahi-machi, Machida, Tokyo 194-8533, Japan (E-mail: snakagawa@xanagen.com, Tel:81-44-829-3031, Fax:81-44-813-1651)  
 COMMENT This sequence is conducted by collaboration of Kyowa Hakko Kogyo Co. Ltd. And Kitasato University.  
 FEATURES  
 source  
 1. 332050  
 /organism="Corynebacterium glutamicum ATCC 13032"  
 /mol\_type="genomic DNA"  
 /strain="ATCC 13032"  
 /db\_xref="taxon:196627"  
 /note="ATCC 13032"  
 314. .2518  
 /gene="Cgl0654"  
 314. .2518  
 /gene="Cgl0654"  
 /note="pF02554:Carbon starvation protein CsrA"

/codon\_start=1  
/transl\_table=1  
/product="Carbon starvation protein, predicted membrane protein"  
/protein\_id="BAB98047.1"  
/db\_xref="GI:21323420"  
/translation="MKKSPKMSPTARVQLLVFGVIAVAGWGAIAFSRGETINSVWLVLAAVGSYIAFSFYARLIEYKVKPKDQATPAFVINDQKDVPTDRVLFGHFAALA  
GAGPLVGPVMAOKMGLPGTLITLIGVIFAGVODYLWVSTRGRGRSGOMVRDM  
GTGCGAGCGLATISIMITIIIVALLIYVNALADSPKCVFSITWTIPIALPBGVYRLR  
RPGKTEVSIIIGVALLAIAGVAVDTSWGEWFTWSKTLTALILGIMAILP  
WLLIAPRDLYSTPMKIGVIGLLAVGLFAPREVQMSVTSFALENGVPVSGLPFP  
LFTIACGASGFHALISSGTPKLVEKESQRMIGYGMIMESFVAMMALITVILID  
RLHFSMNAPLATGDPATPAAEWNSIGLTGADITPEOLSEAESVGESTVIRTCG  
APLAFGMSSETISGFIIGAGAKAWHPALIMEALFITTYDAGTVARFPMDDTLFR  
VPGRRKDSBWTQNNISTYFVCAALGAILIMGVTPPLGINSILPFLFGIANDLLA  
IALVLVVVVKGLYKRAMI PAFLAMDILVTITASMOKI FHSDPALVIAQVANSR  
DASQGLTERGAASPEAIDAVIRNTMIQGILSLFVLVLVVGAAIIVCISKIRAR  
AAGTPLEETPEPTESEFFAFTGFLASSRDXEYQAMWDERPGAPVSSGGH"  
2522. .2725  
/gene="Cg10655"  
2522. .2725  
/gene="Cg10655"  
/codon\_start=1  
/transl\_table=1  
/product="Hypothetical protein"  
/protein\_id="BAB98048.1"  
/db\_xref="GI:21323421"  
/translation="WMALTHALWKIPRAVMWYLTLMGDRAYSRYVHLKHHHDADI  
PTREYVRAKADQDANPGARCC"  
complement(3092. .3703)  
/gene="Cg10656"  
complement(3092. .3703)  
/gene="Cg10656"  
/codon\_start=1  
/transl\_table=1  
/product="Hypothetical protein"  
/protein\_id="BAB98049.1"  
/db\_xref="GI:21323422"  
/translation="MSNVINILYMQAHSHTOVNMLIPSSNATQIATFSSSLCCSQ  
NRNVISVWGVRSEERVMEIKLTHRSVSGSPFRGITSQTEBRCPTTIGMFDIC  
HRTGRANWSTQHSCLHRCIVDDVDFLSVRNLSLWVGNAISDLSOVLVLELFR  
RTVGAHFVNHGMVNTPSORTVNAHTNSLKSNGHCKSCVIL"  
3750. .4706  
/gene="Cg10657"  
3750. .4706  
/gene="Cg10657"  
/codon\_start=1  
/transl\_table=1  
/product="Uncharacterized protein involved in propionate catabolism"  
/protein\_id="BAB98050.1"  
/db\_xref="GI:21323423"  
/translation="MNWTIYQALQHTTATROSKKGAISSWKAPAPAFAGKMSIE  
AYDAMRGEGASPWEGBEDGVIAMLSGDHIYITPLPAGGAKRALITLYTRESHA  
EYOSQAPIDILARSWGEKIAQGLRDVDSVLHSTHTHYIGTGSANDPOKRPDS  
RETIDHSIMYIFAVLEDRAMHERSVAPBAHRRTIELMNKISTVEDEMPRRYS  
VDPKRFKARAVITFRKQTVYBDELAVDADHPGARPPAREQITQKFRILAEVSVE  
KEQDRFLDAQRHLEDELSELNTELDADILAKAPVPEGLF"  
4715. .5629  
/gene="Cg10658"  
4715. .5629  
/gene="Cg10658"  
/codon\_start=1  
/transl\_table=1  
/product="PEP phosphonomutase and related enzymes"  
/protein\_id="BAB98051.1"  
/db\_xref="GI:21323424"  
/translation="MFSSAVAPTERKALAAALAPRIARPGAFSPLAARIOEAF  
EGYVSGAVVAADLALPDIGITTFTEVAHRSQJARTYDLPVLVADADTGEFVSAR  
TVELEDAGVAGCHLEQVNVKRCGHLDGKEVQGTDMVRRIAAVNERDEQVICA  
RTAAAGVEGIDSAIERAKAYADAGADMI FTEALYSPADPEKFAAVDIPILANTEFG  
KTEILPAQLEDIDIGNAVIYPTVLIRAMGQVEQALDIANTGIGTDWVDMQRHSRL

YELLRYNEYNAPDOQVFTYSADSKYPIF"  
5658. .6809  
/gene="Cg10659"  
5658. .6809  
/gene="Cg10659"  
/EC\_number="4.1.3.7"  
/note="PF00285: Citrate synthase"  
/codon\_start=1  
/transl\_table=1  
/product="Citrate synthase"  
/protein\_id="BAB98052.1"  
/db\_xref="GI:21323425"  
/translation="MSSATTTDVKGLYGVIADYAVTSKVPETNSLTYRGVAVEDLV  
EANSFEFVPLHNGELPFAQOAEFNERGSRYSIDAGLISLHSLPKFAHPMDVR  
TAVSWGTQSEKFTPTDSEHIRKVGHTLLAQLPMLVLAODIRRRKGDIIPDSKSYA  
ENLISVYRGTPSSPASNPADVDPEKSLILYVHESFNASTFARVITSTKSYNSAI  
TGAIGLKGPHGCGANEFWHTMLAIDDPKAAWAMNNALDNKQVWVGFRYYKRD  
SRVPSMEKSFRELAARHDEGKWVAMNEMKDADARTGAKPNLDFPAGPYYHLGFPV  
DFFTPLFVIKRVAGWTAHIVEQYENSLIRPLSEYNGEOREVAPIEKR"  
complement(6846. .7040)  
/gene="Cg10660"  
complement(6846. .7040)  
/gene="Cg10660"  
/codon\_start=1  
/transl\_table=1  
/product="Hypothetical protein"  
/protein\_id="BAB98053.1"  
/db\_xref="GI:21323426"  
/translation="MYFDVQDDDELHPSFFELVADQALDAHRCAPHYVWREAVA  
KHVLPQSQKNWVTRMLHNG"  
complement(7129. .8172)  
/gene="Cg10661"  
complement(7129. .8172)  
/gene="Cg10661"  
/EC\_number="1.1.1.82"  
/note="PF02615: Malate/L-lactate dehydrogenase"  
/codon\_start=1  
/transl\_table=1  
/product="Malate/L-lactate dehydrogenases"  
/protein\_id="BAB98054.1"  
/db\_xref="GI:21323427"  
/translation="MPEVYVNAQQLTVLCTDILTKTVGPAADAHLVGDSLVOADLWCH  
PSHGVLRLPMYVRLSHSGAMTTHAHVEVNDLCAVALDGHNGIQVLADHARKAVT  
RAMVFGIAVSVNSNHFGTAMYYTRKAAAGCVSLITTNASPAAMPWGERKIGIN  
PMSTIAPGGETAVDIDIANVARGKIYHARQNNMIPEMTWITSGAFTDPDAIN  
GVYLPMAGHKGYAISPMMDVLSVLTSGSHSTVHSPYPTPGCGHGLFIADVAIF  
RDPQDPDDALSDLVGEVSTPKAQTTEILFYPESEDRARNSAHGISLPEKTMEL  
QELAIENHVYTR"  
complement(8265. .8987)  
/gene="Cg10662"  
complement(8265. .8987)  
/gene="Cg10662"  
/note="PF00392: Bacterial regulatory proteins, gntR family"  
/codon\_start=1  
/transl\_table=1  
/product="Transcriptional regulators"  
/protein\_id="BAB98055.1"  
/db\_xref="GI:21323428"  
/translation="WVQKDAQSPATRKADQVYTQIRREIEDGTLNPGQMSVWLVE  
HTGASRTPVADALRLADELILERQAPMWSPLIRHKDLFEFRRIVEYALBEI  
SVGASRSPRIFGFSTLAADFLELNSADDADTFADFRRTSKSFDLVVANTNOFG  
KSILSLKPHTRRLIYASPHARQSVQEHIMCSAVASGDRSAGACRQHLHVE  
KSILYALINADSGSGGIDIRS"  
complement(9094. .9798)  
/gene="Cg10663"  
complement(9094. .9798)  
/gene="Cg10663"  
/codon\_start=1  
/transl\_table=1  
/product="Hypothetical membrane protein"  
/protein\_id="BAB98056.1"  
/db\_xref="GI:21323429"  
/translation="MRFSRVLPALITTAVSIPTASAAVLTADTDKELCIASNTDOSA

VTTFMNSIDSVREORLDELDAODPEIKAIIESYIAQDNAPPAELQVRLDAISSGE  
GLAMLLPDPPLTADPVAESFKTEYTDKADKIDISFSDSPASDIVLSQCAQATGTR  
TIDAEVADRTDDYNEQSOTALKEDFQCIDAIIDARP1PQYIILIGSALMAVAVL  
GIRAWNSRKQKHSO" .12094)  
/gene="Cg10664"  
/complement (9878..12094)  
/EC number="1.1.1.42"  
/note="TRIGR00178:monomer\_idh: isocitrate dehydrogenase,  
NADP-dependent"  
/codon\_start=1  
/transl\_table=11  
/product="Monomeric isocitrate dehydrogenase (NADP+)"  
/protein\_id="BAB98057.1"  
/db\_xref="GI:21323430"  
/translation="MAKTIWPTDEAPLITATYSLKRVFAFAATACIEVETUJSLAG  
RLAOPPERLTEDQVGNALAEGLBELAPEANILKLPNISASVQLKAIKELDDOG  
YDIPLPDVAITDEDEKDIILARYNAVKGSAVNPEVLRGNSDRAP1AVKNFVQIFPHRM

Query Match 99.4%; Score 2111; DB 1; Length 332050;  
Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 1; Gaps 1;  
Matches 2122; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 CTCGAGCGGAGTGGTGAATCGGCACTCTTAAGCAATGCCGCTTTAAATAAAGCA 60  
Db 213670 CTCGAGCGGAGTGGTGAATCGGCACTCTTAAGCAATGCCGCTTTAAATAAAGCA 213611

Qy 61 CTTAATATGTTCTCACACATCTGGCCGACGACAGAAAGTATGTTGTCATCACAGCTA 120  
Db 213610 CTTAATATGTTCTCACACATCTGGCCGACGACAGAAAGTATGTTGTCATCACAGCTA 213551

Qy 121 AACGTGTAATGTAAGTAACTCACTCAATTCGAAATGCGATGGATTTGAAACATC 180  
Db 213550 AACGTGTAATGTAAGTAACTCACTCAATTCGAAATGCGATGGATTTGAAACATC 213491

Qy 181 ACTCCCCCAATATCTTAACCTTAAGTATGTTTAACTGATCTTAATAAGT 240  
Db 213490 ACTCCCCCAATATCTTAACCTTAAGTATGTTTAACTGATCTTAATAAGT 213431

Qy 241 TCCGATCTACCCCTCTTAAACCCGAAATACCCCTTTTGCAAGATTGCAACACAA 300  
Db 213430 TCCGATCTACCCCTCTTAAACCCGAAATACCCCTTTTGCAAGATTGCAACACAA 213371

Qy 301 GTGCAATGTTAAAGGCTTCAAGTCAACATCTGTCGGTCTTAAAGTATGTTGGG 360  
Db 213370 GTGCAATGTTAAAGGCTTCAAGTCAACATCTGTCGGTCTTAAAGTATGTTGGG 213311

Qy 361 ACGTCTAGGCAAAAAGTATGTTGAGATGAAAGCATATCCGTCATTTTAAAGCA 420  
Db 213310 ACGTCTAGGCAAAAAGTATGTTGAGATGAAAGCATATCCGTCATTTTAAAGCA 213252

Qy 421 TCGATAGCTTAATGAGCTTATGATCTTCCGCTCTAAATAGATAGCAAGACATTG 480  
Db 213251 TCGATAGCTTAATGAGCTTATGATCTTCCGCTCTAAATAGATAGCAAGACATTG 213192

Qy 481 ATTATTAACAAAGCATTTTTCGGCCGCTGAGAAAGCTTTCCGATATGTTGGGCG 540  
Db 213191 ATTATTAACAAAGCATTTTTCGGCCGCTGAGAAAGCTTTTCGATATGTTGGGCG 213132

Qy 541 ATGGAACACATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGAT 600  
Db 213131 ATGGAACACATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGAT 213072

Qy 601 TCTTGAATGAAGCTCCCAATATGACAACTCAATCAAGGCTATGACAGACCTTGA 660  
Db 213071 TCTTGAATGAAGCTCCCAATATGACAACTCAATCAAGGCTATGACAGACCTTGA 213012

Qy 661 CGGGCTGAGAGGAGGCAATGAGATGAATGAGATGAATGAGAGGAGGAGGAGGAG 720  
Db 213011 CGGGCTGAGAGGAGGCAATGAGATGAATGAGATGAATGAGAGGAGGAGGAGGAG 212952

Qy 721 GGCATTCGGGATGCGGTTATTTTGTCCGATTTTCTTCCGCGGAGTTCTTTGGG 780

|||||  
Db 212951 GGCATTCGGGATGCGGTTATTTTGTCCGATTTTCTTCTCCGCGGTTCTTTGGG 212892

Qy 781 GTCCGCTGGTGGGATCATGAAAGCATTCACCGGCCACAGAGTGAATCTCCAC 840  
Db 212891 GTCCGCTGGTGGGATCATGAAAGCATTCACCGGCCACAGAGTGAATCTCCAC 212832

Qy 841 CTGGATCTCCCTGCTTCCGCTGGTGGGATGAGGAGGAGGAGGAGGAGGAGGAG 900  
Db 212831 CTGGATCTCCCTGCTTCCGCTGGTGGGATGAGGAGGAGGAGGAGGAGGAG 212772

Qy 901 ATGATGTCATTAACCGCGCTGTGACGCGTCAACGAGGCGCATTTGCCGTTCTG 960  
Db 212771 ATGATGTCATTAACCGCGCTGTGACGCGTCAACGAGGCGCATTTGCCGTT 212712

Qy 961 TATTGGCGCAATCCACAGATGGTGGGCGGCGCATGCGCTGTGGGATTCATCT 1020  
Db 212711 TATTGGCGCAATCCACAGATGGTGGGCGGCGCATGCGCTGTGGGATTCATCT 212652

Qy 1021 ACTTTTGGGGAACCGGCGCGGATGAGTTTCTGGGCTCTGGGCTGGAGTTAA 1080  
Db 212651 ACTTTTGGGGAACCGGCGCGGATGAGTTTCTGGGCTCTGGGCTGGAGTT 212592

Qy 1081 ACTGGGCAATGCGCTTCCAGACGCTGTGACAGAGCGGAGAAATTTGTAATA 1140  
Db 212591 ACTGGGCAATGCGCTTCCAGACGCTGTGACAGAGCGGAGAAATTTGTAATA 212532

Qy 1141 ATTGATGCAATTTGTCTCCCACTCCCAATTTGCTGAGCGGATGCAAAAACCT 1200  
Db 212531 ATTGATGCAATTTGTCTCCCACTCCCAATTTGCTGAGCGGATGCAAAAACCT 212472

Qy 1201 ATTGACCGGATGAGGAGCAACGAGATGTTTCTCCAAACATCTCGGCTGCA 1260  
Db 212471 ATTGACCGGATGAGGAGCAACGAGATGTTTCTCCAAACATCTCGGCTGCA 212412

Qy 1261 GTGATGAGGCGATTTGCGCTTCTGTGACCGGACGAGGCTGGAATCGGAGAT 1320  
Db 212411 GTGATGAGGCGATTTGCGCTTCTGTGACCGGACGAGGCTGGAATCGGAGAT 212352

Qy 1321 GAAACGTTGGGAGGAGACGTCGTCAAGCTTTCTGAGCGGCTGCGGATGAGCC 1380  
Db 212351 GAAACGTTGGGAGGAGACGTCGTCAAGCTTTCTGAGCGGCTGCGGATGAGCC 212292

Qy 1381 GCTGTGCGGATTTGCGCTGCGGAGATGCGGAGGCGGCGGCTGAGTATGAGGAG 1440  
Db 212291 GCTGTGCGGATTTGCGCTGCGGAGATGCGGAGGCGGCGGCTGAGTATGAGGAG 212232

Qy 1441 CGCGCTTCAAGCTTTGAGCGGACGAGAGCTGCTTTTGGCGCTGCGATTTCCG 1500  
Db 212231 CGCGCTTCAAGCTTTGAGCGGACGAGAGCTGCTTTTGGCGCTGCGATTTCCG 212172

Qy 1501 CGCGAGCTAAACCTGCGATGTCATTCATTCGACACTTCGCGGCGGAAATGTC 1560  
Db 212171 CGCGAGCTAAACCTGCGATGTCATTCATTCGACACTTCGCGGCGGAAATGTC 212112

Qy 1561 GCGGCTGAGAGCTTCGAGATTCGCAAGCTTCATTTGGCGGACCTTGTCCAAGTT 1620  
Db 212111 GCGGCTGAGAGCTTCGAGATTCGCAAGCTTCATTTGGCGGACCTTGTCCAAGTT 212052

Qy 1621 GCTCCCTCCCAACCGTTTCCGATTAATGATGAGGCGTGGGCGGCGCTGAG 1680  
Db 212051 GCTCCCTCCCAACCGTTTCCGATTAATGATGAGGCGTGGGCGGCGCTGAG 211992

Qy 1681 ATGCTGCCGCGATCTGATTCATTCGCGGCGGAAACCGGATGCTCCGATTCG 1740  
Db 211991 ATGCTGCCGCGATCTGATTCATTCGCGGCGGAAACCGGATGCTCCGATTCG 211932

Qy 1741 GAGGCGCTGCGGATCTCTTCCGAGACCAACCAAGCGGCGGAAATATGAGGAG 1800  
Db 211931 GAGGCGCTGCGGATCTCTTCCGAGACCAACCAAGCGGCGGAAATATGAGGAG 211872

Qy 1801 CAAAGCGTGAAGGCGACGCACTTTAAAGCAAGGCTTATGACGAGGATGTCGCG 1860



Dh 211871 CAAAGCGTGCAGGCGCAACGACTTTTAAAGCCAGGGCTTATGACGAGGATGTCGCCGAA 211812  
Qy 1861 ACCGAGACTTGTGTGAAGAAATTCCTGGGCAAACTGAGCAAGCCCTCTCGGAATTGGAT 1920  
Db 211811 ACCGAGACTTGTGTGAAGAAATTCCTGGGCAAACTGAGCAAGCCCTCTCGGAATTGGAT 211752  
Qy 1921 AACAACTCCGAGAGGCGGCGAGCAGTCGCTTCAACAGATTGTGAGCGTTTGAACGCGAG 1980  
Db 211751 AACAACTCCGAGAGGCGGCGAGCAGTCGCTTCAACAGATTGTGAGCGTTTGAACGCGAG 211692  
Qy 1981 TAAAGAAATTTATGCGCTGATCAAAATGATATGAACCAAGGTTACGCGGCAAGACAGTGG 2040  
Db 211691 TAAAGAAATTTATGCGCTGATCAAAATGATATGAACCAAGGTTACGCGGCAAGACAGTGG 211632  
Qy 2041 GTGGCGGGAACCTCAGGGCGGTGAAGCAAGCTCTGGGGGATGTGTCAGCTTGAACGATGCC 2100  
Db 211631 GTGGCGGGAACCTCAGGGCGGTGAAGCAAGCTCTGGGGGATGTGTCAGCTTGAACGATGCC 211572  
Qy 2101 GCCGACCTTCATGCTGGAATTC 2123  
Db 211571 GCCGACCTTCATGCTGGAATTC 211549

RESULT 5  
AX127145/c 349980 bp DNA linear PAT 11-MAY-2001  
LOCUS AX127145  
DEFINITION Sequence 7061 from Patent EP1108790.  
ACCESSION AX127145 AX114121  
VERSION AX127145.1 GI:14041133  
KEYWORDS  
SOURCE Corynebacterium glutamicum  
ORGANISM Corynebacterium glutamicum  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
REFERENCE 1 Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,  
AUTHORS Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.  
TITLE Novel polynucleotides  
JOURNAL Patent: EP 1108790-A 7061 20-JUN-2001;  
KYOMA HAKKO KOGYO CO., LTD. (JP)  
FEATURES  
Source 1. 349980  
Location/Qualifiers  
/organism="Corynebacterium glutamicum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1718"  
/note="Seq 1 to long (3,309,400) split in 11, seq 7061  
0.600,001 0.949,980 ,"  
BASE COUNT 79274 a 90638 c 98727 g 81341 t  
ORIGIN

Query Match 99.4%; Score 2111; DB 6; Length 349980;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2122; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 CTGAGAGGAGAGTGTGATGCGCACTCTTAAGCAATGCGGCTTTAAATTAAGCAA 60  
Db 281620 CTGAGAGGAGAGTGTGATGCGCACTCTTAAGCAATGCGGCTTTAAATTAAGCAA 281561  
Qy 61 CTATATAGTTTCTACACATCTGGCGAGCAACAGATATTTGTCATCAGCTA 120  
Db 281560 CTATATAGTTTCTACACATCTGGCGAGCAACAGATATTTGTCATCAGCTA 281501  
Qy 121 AACGTGTAATGTAGTAACTTAACCTAATGCAATGCGATGCGATTGGAATCTC 180  
Db 281500 AACGTGTAATGTAGTAACTTAACCTAATGCAATGCGATGCGATTGGAATCTC 281441  
Qy 181 ACTCCCCCAATATCTTAACCTTAATTAAGTATGTTTACCTCATTTATTAAGT 240  
Db 281440 ACTCCCCCAATATCTTAACCTTAATTAAGTATGTTTACCTCATTTATTAAGT 281381  
Qy 241 TCCCGATCTACCCCTCTTTAACCAGCAATACCCCTTTTGCAAGATTGCAACACA 300  
Db 281380 TCCCGATCTACCCCTCTTTAACCAGCAATACCCCTTTTGCAAGATTGCAACACA 281321

Qy 301 GTGCAATAGTTAAACGGGCTTCAACGTCACCAATTTCTCGGTTTAAAGGATGTTGGG 360  
Db 281320 GTGCAATAGTTAAACGGGCTTCAACGTCACCAATTTCTCGGTTTAAAGGATGTTGGG 281261  
Qy 361 ACGTCTAGGCAAAAGTATGTTGTAGATGAAGCAATATCCGTCATTTTAAACGCA 420  
Db 281260 ACGTCTAGGCAAAAGTATGTTGTAGATGAAGCAATATCCGTCATTTTAAACGCA 281202  
Qy 421 TCGATACCTTAATTTGGGCTTAATCTTCCGCTTAATTAAGTATGCAAGACATTGCA 480  
Db 281201 TCGATACCTTAATTTGGGCTTAATCTTCCGCTTAATTAAGTATGCAAGACATTGCA 281142  
Qy 481 ATTAATTAACAAAGCCATTTTGGGCGTGAAGAGGTTTCCGATATAGTGTGGGGC 540  
Db 281141 ATTAATTAACAAAGCCATTTTGGGCGTGAAGAGGTTTCCGATATAGTGTGGGGC 281082  
Qy 541 ATGAAACACACTTCAGATTGACGCTCATATGACTGGTTTGAACCTGACAGCTTCATT 600  
Db 281081 ATGAAACACACTTCAGATTGACGCTCATATGACTGGTTTGAACCTGACAGCTTCATT 281022  
Qy 601 TCTTGAATGAATCTCCCAATATGACAACTCATATCAAGGCTATGACAGACCTTGAG 660  
Db 281021 TCTTGAATGAATCTCCCAATATGACAACTCATATCAAGGCTATGACAGACCTTGAG 280962  
Qy 661 CGGGCTGAAGCAAGGCAAAATGAGATGATCGGTATTAATCTGGAAGAGGACCGTGGAG 720  
Db 280961 CGGGCTGAAGCAAGGCAAAATGAGATGATCGGTATTAATCTGGAAGAGGACCGTGGAG 280902  
Qy 721 GGCATTCGGGTAGCCGTTATTTTGTCCGATTTTCTCTCGGCGGTTCTTTGGGACG 780  
Db 280901 GGCATTCGGGTAGCCGTTATTTTGTCCGATTTTCTCTCGGCGGTTCTTTGGGACG 280842  
Qy 781 GTGGCGTGGTGGCATATGAAGGCGATTCACCGGCGCAAGAGCTGAAATCTCCACTG 840  
Db 280841 GTGGCGTGGTGGCATATGAAGGCGATTCACCGGCGCAAGAGCTGAAATCTCCACTG 280782  
Qy 841 CTGGTCTCCCTGCTTCCGGTGTGGCGGAGTGCAGGAAGAACAATGAGCTTTGTGATG 900  
Db 280781 CTGGTCTCCCTGCTTCCGGTGTGGCGGAGTGCAGGAAGAACAATGAGCTTTGTGATG 280722  
Qy 901 ATGATGTCATTAACCGGCGCTGTGCAGCGTCAACCGGAGGCGATTTCCGTTCTGGTG 960  
Db 280721 ATGATGTCATTAACCGGCGCTGTGTGCAGCGTCAACCGGAGGCGATTTCCGTTCTGGTG 280662  
Qy 961 TATTTGGCAATCCCAAGATGGGTGGCGCATAGGCTCTGTGGGTTTATTTGGGCAATCTC 1020  
Db 280661 TATTTGGCAATCCCAAGATGGGTGGCGCATAGGCTCTGTGGGTTTATTTGGGCAATCTC 280602  
Qy 1021 ACTTTTGGGAAACCGGCGCGGAGATAGTTTCTGAGTCTGCGGTGGTGAAGTAAAC 1080  
Db 280601 ACTTTTGGGAAACCGGCGCGGAGATAGTTTCTGAGTCTGCGGTGGTGAAGTAAAC 280542  
Qy 1081 ACTGGGATGCGCTTCCAGCGGTGTGCAGCAGGCGGAATTTGTGTAATCTGTTG 1140  
Db 280541 ACTGGGATGCGCTTCCAGCGGTGTGCAGCAGGCGGAATTTGTGTAATCTGTTG 280482  
Qy 1141 ATTGATGAATTTGTGCGCACTTCAATGCGTGTGCAAGCGGTGCAAAAACCTCAAGTT 1200  
Db 280481 ATTGATGAATTTGTGCGCACTTCAATGCGTGTGCAAGCGGTGCAAAAACCTCAAGTT 280422  
Qy 1201 ATTACGCGGTAGAGCAACGATGCTTTTCTCCAAACAATCTGCGCGGTGCACTTCCG 1260  
Db 280421 ATTACGCGGTAGAGCAACGATGCTTTTCTCCAAACAATCTGCGCGGTGCACTTCCG 280362  
Qy 1261 GTGATGAGCGAATGCGGTTCTCGTGAACCGCAAGGCGTGAATTCGGGAGATTAAG 1320  
Db 280361 GTGATGAGCGAATGCGGTTCTCGTGAACCGCAAGGCGTGAATTCGGGAGATTAAG 280302  
Qy 1321 GAAACGTTGGGGCAAGCGTGTCAAGCTTTCTGTGCGCGTGTGCGGATTTGAGCCG 1380  
Db 280301 GAAACGTTGGGGCAAGCGTGTCAAGCTTTCTGTGCGCGTGTGCGGATTTGAGCCG 280242





```

|||||
Db 721 AATTGTGTCCGACATCCAAATTGGCGAGCGGTGGCAAAAACCTCAAGATTATTACGCC 780
|||
Qy 1209 GGTGAGGCAACGGAATGTTTTTCTCCAAACAATCTTGCGGTGGCACTTCCGGTGAATGGA 1268
|||
Db 781 GGTGAGGCAACGGAATGTTTTTCTCCAAACAATCTTGCGGTGGCACTTCCGGTGAATGGA 840
|||
Qy 1269 GGGGATTTGGCGGCTCTGTGTAACCGGAGAGGCTGGGAATGGGGGATTTATGAAAGCTT 1328
|||
Db 841 GGGGATTTGGCGGCTCTGTGTAACCGGAGAGGCTGGGAATGGGGGATTTATGAAAGCTT 900
|||
Qy 1329 GGGGGAGAGCTGTCTCAAGCTTTCTGTGCGCGTGTGCGGCATTTAGAGCCCGGCTGTGCG 1388
|||
Db 901 GGGGGAGAGCTGTCTCAAGCTTTCTGTGCGCGTGTGCGGCATTTAGAGCCCGGCTGTGCG 960
|||
Qy 1389 GGTGCGCTGTGCGGCATCGGGGGGCGGCGCGTGTGCTGATTTGGGAGAGATCCCGCTT 1448
|||
Db 961 GGTGCGCTGTGCGGCATCGGGGGGCGGCGCGTGTGCTGATTTGGGAGAGATCCCGCTT 1020
|||
Qy 1449 CAGGCTTGGGCGGAGAGAGCTGGGCTTGGCGGTGTGAGCATTTGCGTGGGCGCGAGACT 1508
|||
Db 1021 CAGGCTTGGGCGGAGAGAGCTGGGCTTGGCGGTGTGAGCATTTGCGTGGGCGCGAGACT 1080
|||
Qy 1509 AAACCTGCGGATCGTGTCCATCATCGACACCTCGGCGCGCAATTTGTCAGAGCGGCTGA 1568
|||
Db 1081 AAACCTGCGGATCGTGTCCATCATCGACACCTCGGCGCGCAATTTGTCAGAGCGGCTGA 1140
|||
Qy 1569 GAGAGCTGGGCAATCGCAAGCTGATGTCGGGCACTTGTCCAACTTATCGACCTCCCTT 1628
|||
Db 1141 GAGAGCTGGGCAATCGCAAGCTGATGTCGGGCACTTGTCCAACTTATCGACCTCCCTT 1200
|||
Qy 1629 CCGGACGCTTGGGCAATTTATGTCAGGGGCTTGGGCGGTGGCGGCTGGGCGATGCTGCC 1688
|||
Db 1201 CCGGACGCTTGGGCAATTTATGTCAGGGGCTTGGGCGGTGGCGGCTGGGCGATGCTGCC 1260
|||
Qy 1689 CGCGGATCTGTCTACGCGGCGGCAAAAACGCGTGTGCTGTCCGATTTGCCACAGAGGCGC 1748
|||
Db 1261 CGCGGATCTGTCTACGCGGCGGCAAAAACGCGTGTGCTGTCCGATTTGCCACAGAGGCGC 1320
|||
Qy 1749 CTCGGGCAATCTCTTCCGCGCAACCAACGCGCGGGAATCATATGAGCAAGAGCGT 1808
|||
Db 1321 CTCGGGCAATCTCTTCCGCGCAACCAACGCGCGGGAATCATATGAGCAAGAGCGT 1380
|||
Qy 1809 GCAGGCGGAGCACTTTTAAGCAAGGCTTATGAGCGGATCGTCCGCGAAACGAGCA 1868
|||
Db 1381 GCAGGCGGAGCACTTTTAAGCAAGGCTTATGAGCGGATCGTCCGCGAAACGAGCA 1440
|||
Qy 1869 CTTTGTGAAGAAATTCGCGCAATCAGCAAGCCCTCTCCGAATTTGATTAACAATCC 1928
|||
Db 1441 CTTTGTGAAGAAATTCGCGCAATCAGCAAGCCCTCTCCGAATTTGATTAACAATCC 1500
|||
Qy 1929 GAGAGGCGGAGCGGAGCGGACAGTGCCTTCAACGATTTGAGCGTTTAAAGCAAGAA 1988
|||
Db 1501 GAGAGGCGGAGCGGAGCGGACAGTGCCTTCAACGATTTGAGCGTTTAAAGCAAGAA 1560
|||
Qy 1989 ATTATGCGCTGATCA 2003
|||
Db 1561 ATTATGCGCTGATCA 1575
|||

```

```

RESULT 7
AR202320 1473 bp DNA linear PAT 20-APR-2002
LOCUS AR202320
DEFINITION Sequence 2 from patent US 6361986.
ACCESSION AR202320
VERSION AR202320.1 GI:20256859
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1473)
AUTHORS Tilly, Y., Eikmanns, B., Eggeling, L., Sahn, H. and Mockel, B.
TITLE Process for the preparation of L-amino acids by fermentation and

```

```

JOURNAL. nucleotide sequences coding for the accD gene
Patent: US 6361986-A 2 26-MAR-2002;
FEATURES
Source Location/Qualifiers
1..1473
BASE COUNT 275 a 414 c 456 g 328 t
ORIGIN
Query Match 69.4%; Score 1473; DB 6; Length 1473;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1473; Conservat 0; Mismatches 0; Indels 0; Gaps 0;
508 GTGAGAGAGGCTTTTCGACTATGATGTGGGCGATGGAACACACTTCAGATTGACGCTC 567
1 GTGAGAGAGGCTTTTCGACTATGATGTGGGCGATGGAACACACTTCAGATTGACGCTC 60
568 ATAGACTCGGTTTGAACCTCGACAGCTTCAATTTCTTGAATGAACCTCCCAATATGAC 627
61 ATAGACTCGGTTTGAACCTCGACAGCTTCAATTTCTTGAATGAACCTCCCAATATGAC 120
628 AACCTCAATCAAGGCTATGAGAGACCTTGGACCGGCTCGAAGCAAGGCCAATGAGAT 687
121 AACCTCAATCAAGGCTATGAGAGACCTTGGACCGGCTCGAAGCAAGGCCAATGAGAT 180
688 GAATCGGTAATTACTGGAGAAAGGACCGTGAAGGCAATTCGCGTAAAGCGTTATTTTGTCC 747
181 GAATCGGTAATTACTGGAGAAAGGACCGTGAAGGCAATTCGCGTAAAGCGTTATTTTGTCC 240
748 GATTTTCTTCTCGGCGGCTTTTGGGACGCGTGTGCGATCATGAAGGCG 807
241 GATTTTCTTCTCGGCGGCTTTTGGGACGCGTGTGCGATCATGAAGGCG 300
808 ATTACCGGCGCCACAGAGCTGAAACCTCCACCTGTCTCCCTGCTTCGGTGGTGGC 867
301 ATTACCGGCGCCACAGAGCTGAAACCTCCACCTGTCTCCCTGCTTCGGTGGTGGC 360
868 CGCATGAGAGAAACATCGAGCTTTTGTATGATGATGTCATTAACCGGCGCTGTGAC 927
361 CGCATGAGAGAAACATCGAGCTTTTGTATGATGATGTCATTAACCGGCGCTGTGAC 420
928 CGTACCGGAGGCGCAATTTGACGCTTCTGTGTATTTGGCAATCCAGATGGGTGGC 987
421 CGTACCGGAGGCGCAATTTGACGCTTCTGTGTATTTGGCAATCCAGATGGGTGGC 480
988 GCCATGCGCTCTGTGGGTTATCTGGGCAATCTCACTTTTGGGAAACCGGCGGCGAGATA 1047
481 GCCATGCGCTCTGTGGGTTATCTGGGCAATCTCACTTTTGGGAAACCGGCGGCGAGATA 540
1048 GGTTCCTGAGGCTCTGCGGTGTGAGTTAACCACTGGGCGATGCGCTTCAGACGAGTGTG 1107
541 GGTTCCTGAGGCTCTGCGGTGTGAGTTAACCACTGGGCGATGCGCTTCAGACGAGTGTG 600
1108 CAGCAGCGGAGAAATTTGTAAGAACTGATGTATGATGAATTTGTGCGCACTTCAA 1167
601 CAGCAGCGGAGAAATTTGTAAGAACTGATGTATGATGAATTTGTGCGCACTTCAA 660
1168 TTGCGTGCACGCGTGGGAAACCTTCAGAGTTATTCAGCCGCTGAGAGCAAGGATCGT 1227
661 TTGCGTGCACGCGTGGGAAACCTTCAGAGTTATTCAGCCGCTGAGAGCAAGGATCGT 720
1228 TTTTCTGCAACACTCCGTGGCGTGGCACTTCGCGGTGATGGAGGCGATTTGCGCTTCTGCT 1287
721 TTTTCTGCAACACTCCGTGGCGTGGCACTTCGCGGTGATGGAGGCGATTTGCGCTTCTGCT 780
1288 GACCCGAGAGGCTGTGAATCGGAGAGATTATGAAACGTTGGGCGGAGACGTCGTAAG 1347
781 GACCCGAGAGGCTGTGAATCGGAGAGATTATGAAACGTTGGGCGGAGACGTCGTAAG 840
1348 CTTTCTGTGTCGCGTGTGCGCATTTGAGCCGCGCTGTGCGCTTCTGCGCGCATC 1407
841 CTTTCTGTGTCGCGTGTGCGCATTTGAGCCGCGCTGTGCGCTTCTGCGCGCATC 900
1408 GGGGGCGGCGCGGTGTGTCGATTTGGGCAAGGATGCGCGCTTCAAGCTTGGGCGGAGAG 1467

```

Db	901	GGGGGGCCGGCCCGTGTGTGTGATTGGGCGAGATCGCCGCTTACCGCTTGGGCGCGAGAG	960
Qy	1468	CTGCCTTTTGGCGGTCGTGTGGCATTTTCGTGGCGCGCGAGCTAAACCTGCCATCGTGTCC	1522
Db	961	CTGGCTTTTGGCGCGTGTGTGGCATTTTCGTGGCGCGCGAGCTAAACCTGCCATCGTGTCC	1022
Qy	1528	ATCATCGACACTCCGGCGCGGAATTTGTGCAAGGCGGCTGAGGAGCTCGGCATCGCAAGC	1587
Db	1021	ATCATCGACACTCCGGCGCGGAATTTGTGCAAGGCGGCTGAGGAGCTCGGCATCGCAAGC	1080
Qy	1588	TGATTTGGCGGCACTTTGTCCAAGCTTATCGACGCTCCCTTCCACCGTTTCGGTCATT	1647
Db	1081	TGATTTGGCGGCACTTTGTCCAAGCTTATCGACGCTCCCTTCCACCGTTTCGGTCATT	1140
Qy	1648	ATTGGTCAAGGGCGTTGGGCGGTGGCGCGCGTCGTCGTCGCGCGGATCTGGTCTACACCG	1707
Db	1141	ATTGGTCAAGGGCGTTGGGCGGTGGCGCGCGTCGTCGTCGCGCGGATCTGGTCTACACCG	1200
Qy	1708	GCCGAAAAAGCGTGTGCTGTCCGCATTTGGCCACAGAGAGGCGGCTCGGCATCTCTTCCGG	1767
Db	1201	GCCGAAAAAGCGTGTGCTGTCCGCATTTGGCCACAGAGAGGCGGCTCGGCATCTCTTCCGG	1266
Qy	1768	GACACCAACCAACGCCGCGGAAATCATTAAGACGACAAAGGCTGTACAGCGCAGCACTTTTA	1827
Db	1261	GACACCAACCAACGCCGCGGAAATCATTAAGACGACAAAGGCTGTACAGCGCAGCACTTTTA	1322
Qy	1828	AGCCAAAGGCTTATTCAGCGGATGCTCGCCGAAACCGAGCACTTTGTGAAGAAATTCG	1887
Db	1321	AGCCAAAGGCTTATTCAGCGGATGCTCGCCGAAACCGAGCACTTTGTGAAGAAATTCG	1380
Qy	1888	GGCACAATTCAGCAAGCCCTCTCCGAATTTGATTAACAATCCGGAAGAGGCGGGACGCGAC	1947
Db	1381	GGCACAATTCAGCAAGCCCTCTCCGAATTTGATTAACAATCCGGAAGAGGCGGGACGCGAC	1440
Qy	1948	AGTGCCTTCACACGATTTGAGCGTTTAAGCGGAG	1980
Db	1441	AGTGCCTTCACACGATTTGAGCGTTTAAGCGGAG	1473

	RESULT 8			
LOCUS	AX121013			
DEFINITION	Sequence 929 from Patent EP1108790.	1473 bp	DNA	linear
ACCESSION	AX121013			
VERSION	AX121013.1	GI:14037728		
KEYWORDS				
SOURCE				
ORGANISM	Corynebacterium glutamicum Corynebacterium glutamicum Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium.			
REFERENCE	1			
AUTHORS	Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K., Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A. Novel polynucleotides Patent: EP 1108790-A 929 20-JUN-2001; KIYOMA HAKKO KOGYO CO., LTD. (JP)			
TITLE				
JOURNAL				
FEATURES	location/Qualifiers			
source	1..1473 /organism="Corynebacterium glutamicum" /mol_type="genomic DNA" /db_xref="taxon:1718"			
BASE COUNT	275 a     414 c     456 g     328 t			
ORIGIN				

```

Query Match 69.4%; Score 1473; DB 6; Length 1473;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 GTGGAGAGCGTTTCCGACTATGCTGTGGGGCATGAAACACACTTCAGATTGACGCTC 567
|||||
Db 1 GTGGAGAGCGTTTCCGACTATGCTGTGGGGCATGAAACACACTTCAGATTGACGCTC 60

```

[illegible]

|||||  
Db 1141 ATTGTCAGGGCGTTGGCGGTGGCGCGCTGSCATGCTCCGCGATCTGTCTACGCG 1200  
Qy 1708 GCCGAAAACGGGTGGTCTCCGCAATTGCAACAGAGGCGCTCGGCATCTTCCGC 1767  
Db 1201 GCCGAAAACGGGTGGTCTCCGCAATTGCAACAGAGGCGCTCGGCATCTTCCGC 1260  
Qy 1768 GACACCAACGACGCGCGGAAATCATAGAGCAAGGCGTGACGCGACGCACTTTTA 1827  
Db 1261 GACACCAACGACGCGCGGAAATCATAGAGCAAGGCGTGACGCGACGCACTTTTA 1320  
Qy 1828 AGCCAGGGCTTATCGACGCGATCGTCGCGGAAACGAGCACTTTGTAAGAAATTC 1887  
Db 1321 AGCCAGGGCTTATCGACGCGATCGTCGCGGAAACGAGCACTTTGTAAGAAATTC 1380  
Qy 1888 GGCAATTCAGCAACGCGCTCTCCGAATGATTAACATCCGGAGAGGCGCGGAC 1947  
Db 1381 GGCAATTCAGCAACGCGCTCTCCGAATGATTAACATCCGGAGAGGCGCGGAC 1440  
Qy 1948 AGTCGCTTACACGATTTGAGCGTTTACGCGAG 1980  
Db 1441 AGTCGCTTACACGATTTGAGCGTTTACGCGAG 1473

RESULT 9  
BD004742 1473 bp DNA linear PAT 31-JAN-2002  
LOCUS  
DEFINITION Replicable DNA, amino acid sequence, Corynebacterium microorganism,  
shuttle vector, and process for producing L-amino acid.  
BD004742  
ACCESSION BD004742.1 GI:18632703  
VERSION JP 2001008693-A/2.  
KEYWORDS  
SOURCE Corynebacterium glutamicum  
ORGANISM Corynebacterium glutamicum  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Corynebacteriaceae; Corynebacterium.

REFERENCE  
AUTHORS 1. Eggering, U., Eickmans, B., Zamu, H. and Meck, V.  
TITLE Replicable DNA, amino acid sequence, Corynebacterium microorganism,  
shuttle vector, and process for producing L-amino acid  
JOURNAL Patent: JP 2001008693-A 2 16-JAN-2001;  
DEBUSA HUELS AG,FORSCHUNGSZENTRUM JUELICH GMBH  
COMMENT OS Corynebacterium glutamicum  
PN JP 2001008693-A/2

PD 16-JAN-2001  
PR 24-MAY-2000 JP 2000153547  
PI 27-MAY-1999 DE 19924365.4  
PI IVERNE TILKE,LOTHAR EGGERING,BERNHARD EICKMANS,HERMANN ZAMU,  
PI VETTYNA MECKEL  
PC C12N15/09,C12N1/21,C12P13/04,C12P13/06,C12P13/08,PC  
C12P13/08

PC C12P13/12,C12P13/20,C12P21/02//C12N15/09,C12R1/15),(C12N1/21,  
PC C12R1/15),  
PC (C12P13/04,C12R1/15),(C12P13/06,C12R1/15),(C12P13/08,C12R1/15),PC  
(C12P13/12,C12R1/15),(C12P13/20,C12R1/15),(C12P21/02,C12R1/15),PC  
C12N15/00,C12R1/15)

FEATURES  
source 1.1473 Location/Qualifiers  
FT Key (1). (1473).  
CDS Location/Qualifiers

BASE COUNT 275 a 414 c 456 g 328 t  
ORIGIN

Query Match 69.4%; Score 1473; DB 6; Length 1473;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 508 GTGAGAGCGTTTTCGACTATGCTGTGGGCGCATGGAACAACATTCAGATTGACGCTC 567  
Db 1 GTGAGAGAGGTTTTCGACTATGCTGTGGGCGCATGGAACAACATTCAGATTGACGCTC 60  
Qy 568 ATGACTCGGTTTGGACCTCTGACAGTTTATTTTGGAAATGAAATCTCCCAATATGAC 627  
Db 61 ATGACTCGGTTTGGACCTCTGACAGTTTATTTTGGAAATGAAATCTCCCAATATGAC 120  
Qy 628 AACCTCAATGAGGCTATGAGAGACCTTGGAGGCGCTGAAAGCAAGGCAATGCGAT 687  
Db 121 AACCTCAATGAGGCTATGAGAGACCTTGGAGGCGCTGAAAGCAAGGCAATGCGAT 180  
Qy 688 GAATCGTAATTAATGAGAGGACCGTGAAGGCAATTCGCGTATGTTATTTGTC 747  
Db 181 GAATCGTAATTAATGAGAGGACCGTGAAGGCAATTCGCGTATGTTATTTGTC 240  
Qy 748 GATTTTCTCTCTCGGCGGTTCTTTGGGACGCTCGCGTGGGCAATGAAAGCG 807  
Db 241 GATTTTCTCTCTCGGCGGTTCTTTGGGACGCTCGCGTGGGCAATGAAAGCG 300  
Qy 808 ATTGACCGGCGACAGAGCTGAAACTCCACTGCTGCTCCCTGCTCCGCGTGGTGC 867  
Db 301 ATTGACCGGCGACAGAGCTGAAACTCCACTGCTGCTCCCTGCTCCGCGTGGTGC 360  
Qy 868 CGCATGAGAGAGCAATCGAGCTTTTGTGATGATGATGATGATGATGATGATGATGAT 927  
Db 361 CGCATGAGAGAGCAATCGAGCTTTTGTGATGATGATGATGATGATGATGATGATGAT 420  
Qy 928 CGTCAACGCGAGGCGCATTTGCGGTTCTGTGTATTTGGCAATCCGAGATGAGTGGC 987  
Db 421 CGTCAACGCGAGGCGCATTTGCGGTTCTGTGTATTTGGCAATCCGAGATGAGTGGC 480  
Qy 988 GCCATGAGCTCGTGGGCTTCATCGGAGCTGAGCACTTTGCGGAAACCGGCGCGCAGATA 1047  
Db 481 GCCATGAGCTCGTGGGCTTCATCGGAGCTGAGCACTTTGCGGAAACCGGCGCGCAGATA 540  
Qy 1048 GGTTCCTGGGCTCTCGCGTGTGAGATTAACAATGAGGATGCGCTTCAGAGGATGTG 1107  
Db 541 GGTTCCTGGGCTCTCGCGTGTGAGATTAACAATGAGGATGCGCTTCAGAGGATGTG 600  
Qy 1108 CAGCAGCGGAGAAATTTGGTGAACCTGTGTGATGATGATGATGATGATGATGATGAT 1167  
Db 601 CAGCAGCGGAGAAATTTGGTGAACCTGTGTGATGATGATGATGATGATGATGATGAT 660  
Qy 1168 TTGCGTGCAGCGGTGGGCAAAACCTCAAGGTTATTAGCGGATGAGGCAAGGATCGT 1227  
Db 661 TTGCGTGCAGCGGTGGGCAAAACCTCAAGGTTATTAGCGGATGAGGCAAGGATCGT 720  
Qy 1228 TTTTCTCCAACTCTCGCGTGTGCACTTCGCGTGTGATGAGGCGATTCGCTCGT 1287  
Db 721 TTTTCTCCAACTCTCGCGTGTGCACTTCGCGTGTGATGAGGCGATTCGCTCGT 780  
Qy 1288 GACCGGAGAGGCTGGAATGCGGAGATTTATGAAAAGTTGGGGGCAAGCTGTCAAG 1347  
Db 781 GACCGGAGAGGCTGGAATGCGGAGATTTATGAAAAGTTGGGGGCAAGCTGTCAAG 840  
Qy 1348 CTTTCTGTGTGCGTGTGCGGCAATTGAGCCCGGCTGTGGCGTTCGCGGCGGATC 1407  
Db 841 CTTTCTGTGTGCGTGTGCGGCAATTGAGCCCGGCTGTGGCGTTCGCGGCGGATC 900  
Qy 1408 GGGGCGCGCGCGGTGTGCTGATTTGGCAGAGATGCGCTTCAAGCTTGGGCGCAGAG 1467  
Db 901 GGGGCGCGCGCGGTGTGCTGATTTGGCAGAGATGCGCTTCAAGCTTGGGCGCAGAG 960  
Qy 1468 CTGCGTTTGTGCGGTGTGCAATTTGCTGTGCGGCGGCAAGCTTGGCGATGTGTC 1527  
Db 961 CTGCGTTTGTGCGGTGTGCAATTTGCTGTGCGGCGGCAAGCTTGGCGATGTGTC 1020  
Qy 1528 ATCATGACACTTCGCGCGCGCAATTTGTGCAAGCGGCTGAGAGCTCGCATGCGCAAGC 1587  
Db 1021 ATCATGACACTTCGCGCGCGCAATTTGTGCAAGCGGCTGAGAGCTCGCATGCGCAAGC 1080  
Qy 1588 TCGATTTGGGCGCACTTGTCAAGCTTATGACGCTCCCTCCCAAGCGTTTCGTCATT 1647

Db	1081	TCGATTTGGCGCGCACTTGTGCCAGTTTACAGCGCTCCCTTCGCCACCGTTTGGGTCANT	1140
OY	1648	ATTGTGTCAGGGCGTTGGCGGTGGCGCGCTGGCCATGTCCCGCCGCATCTGGTCTACACGG	1707
Db	1141	ATTGTGTACGGGGGTGTGGCGGTGGCGCGCTGGCGCATGTCTGGCCCGGCATCTGGTCTACACGG	1207
OY	1708	GCCGAAAAACGGGTGGCTGTCTCCGCATTTGCCACAGAGGGCGCCTCGCGCCATCTCTTCCGG	1767
Db	1201	GCCGAAAAACGGGTGGCTGTCTCCGCATTTGCCACAGAGGGGGCCTGGGCCATCTCTTCCGG	1267
OY	1768	GACACCAACCAACGCCCGCGGAAATCATAGACGACAAAGCGTGGACGGCGCACGCACTTTTA	1827
Db	1261	GACACCAACCAACGCCCGCGGAAATCATAGACGACAAAGGGGTGGACGGCGCACGCACTTTTA	1327
OY	1828	AGCCCAAGGGCTTATTGACAGGGATGTGTGGCCGAACCGACACTTTGTTGAAAGAAATTTTC	1887
Db	1321	AGCCCAAGGGCTTATTGACAGGGATGTGTGGCCGAACCGACACTTTGTTGAAAGAAATTTTC	1387
OY	1888	GGCACAATCAGACCAACGCCCTCTCCGAATTTGATATACATCCGGAGAGGGCGGAGACGCAC	1947
Db	1381	GGCACAATCAGACCAACGCCCTCTCCGAATTTGATATACATCCGGAGAGGGCGGAGACGCAC	1447
OY	1948	AGTCGCTTCAACAGCAATTTGAGGGTTAGCGCAG	1980
Db	1441	AGTCGCTTCAACAGCAATTTGAGGGTTAGCGCAG	1473

RESULT 10			
BD163130			
LOCUS	BD163130	1473 bp	DNA
DEFINITION	Novel polynucleotide.		
ACCESSION	BD163130		linear
VERSION	BD163130.1	GI:27868892	
KEYWORDS	JP 2002191370-A/929.		
SOURCE	unidentified		

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 1473)	Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ohishi, K., Yocoi, H., Tateishi, N., Senoo, A., Ikeda, M. and Ozaki, A.	Novel polymycolide Patent: JP 2002191370-A 929 09-JUL-2002;		
	KIOWA HAKKO KOGYO CO LTD			
	OS			
	Corynebacterium glutamicum			

PF 15-DEC-2000 JP 20000405096  
PI SATOSHI NAKAGAWA, HIROSHI MIZOGUCHI, SEIKO ANDO, MIKIO HAYASHI,  
PI KEIJO OCHIAI,  
PI HARUHIKO YOKOI, NAOKO TATEISHI, AKIHIRO SENOO, MASATO IKEDA, AKIO  
PI OZAKI  
PC C12N15/09, C12N15/09, C07K14/34, C07K16/12, C07K16/40, C12M1/00, PC  
C12M1/15,  
PC C12N1/19, C12N1/21, C12N5/10, C12N9/00, C12N9/02, C12P7/40, C12P13/  
PC 04, C12P13/08,  
PC C12P19/00, C12P19/34, C12P21/02, C12Q1/37, C12Q1/68, G01N33/53, PC  
G01N33/566,  
PC G01N33/569, G01N33/68, G01N37/00//C12P21/08, (C12N1/21, C12R1:15),  
PC (C12N1/21, C12R1:13), (C12M1/21, C12R1:01), (C12P13/08, C12R1:15),  
PC C12N15/00, C12N15/00  
CC Novel polynucleotide  
FH Key Location/Qualifiers  
FT source 1.1473  
/organism='Corynebacterium glutamicum'.

FEATURES	source	Location/Qualifiers
BASE COUNT	275	a 414 c 456 g 328 t
ORIGIN		

Query Match	69.4%	Score 1473;	DB 6;	Length 1473;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1473; Conservative	0;	Mismatches	0;	Indels 0;
		Gaps		0

QY	508	TTGGAGGAGCGCTTTTCCGACTATGGTGTGGGGCATGGAACAACATTCACAGCATTTGACGCTC	567
Db	1	GTGGAGGAAGCGTTTTCGACTATGGTGTGGGGCATGGAACAACATTCACAGCATTTGACGCTC	60
QY	568	ATGAGCTCGGTTTTGAGCCCTGACAGCTTCATTTCTTGGAATGAATCTCCCAATATGAC	627
Db	61	ATGAGCTCGGTTTTGAGCCCTGACAGCTTCATTTCTTGGAATGAATCTCCCAATATGAC	120
QY	628	AACCGCATCAAGCGTATGACAGACCTTTGAGCGGGCTCCAGGAAGGCCAAATGCCAT	687
Db	121	AACCGCATCAAGCGTATGACAGACCTTTGAGCGGGCTCCAGGAAGGCCAAATGCCAT	180
QY	688	GAATCGGTAATTACTGAGGAAGGACCGTGAAGGGCATTTCCGGTAGCGTTATTTGTCC	747
Db	181	GAATCGGTAATTACTGAGGAAGGACCGTGAAGGGCATTTCCGGTAGCGTTATTTGTCC	240
QY	748	GATTTTTCCTTCCTCGCGGCTTCTTTGGGCACGGTCGGTCGGTCGATCATGAAGCG	807
Db	241	GATTTTTCCTTCCTCGCGGCTTCTTTGGGCACGGTCGGTCGGTCGATCATGAAGCG	300
QY	808	ATTACCGCGGCACAGAGCTGAATCCCACTGCGTGTCTCCCTGCTTCGGTGGTTCG	867
Db	301	ATTACCGCGGCACAGAGCTGAATCCCACTGCGTGTCTCCCTGCTTCGGTGGTTCG	360
QY	868	CGCATGCAAGAGAACAATCGAGCTTTGTGCATGATNGGTGCATTAACCGCGGCTGTGAG	927
Db	361	CGCATGCAAGAGAACAATCGAGCTTTGTGCATGATNGGTGCATTAACCGCGGCTGTGAG	420
QY	928	CGTACCGCGCAGGCGGCATTTGCTCCGTTCTCGGTGATTTTGGCAATCCACAGATGGTGGC	987
Db	421	CGTACCGCGCAGGCGGCATTTGCTCCGTTCTCGGTGATTTTGGCAATCCACAGATGGTGGC	480
QY	988	GCCATNGCGCTGTGAGGGGTTCATCTGGGGCATCTTTGGGGAACCGGCGCGCAGATA	1047
Db	481	GCCATNGCGCTGTGAGGGGTTCATCTGGGGCATCTTTGGGGAACCGGCGCGCAGATA	540
QY	1048	GGTTTCCGTGGTCCCTCGCGTGTGAGGATTAACAATGGGCATGTGGCTTCCAGACGATGTG	1107
Db	541	GGTTTCCGTGGTCCCTCGCGTGTGAGGATTAACAATGGGCATGTGGCTTCCAGACGATGTG	600
QY	1108	CAGCAGCGCGAGAAATTTGTGTGAATACTGTGTGATTAATGGAATTTGTGTGCTCCACTCCAA	1167
Db	601	CAGCAGCGCGAGAAATTTGTGTGAATACTGTGTGATTAATGGAATTTGTGTGCTCCACTCCAA	660
QY	1168	TTGGGTGACGGGTGGGGAATAACCTCAAGGTTATTCAGCCGGTATGAGGCAACGATTCGT	1227
Db	661	TTGGGTGACGGGTGGGGAATAACCTCAAGGTTATTCAGCCGGTATGAGGCAACGATTCGT	720
QY	1228	TTTTCTTCCAAACAATCTTCGCGTGTGGCACTTCGCGTGAATGAGAGCGAATTCGCGCTTTCGT	1287
Db	721	TTTTCTTCCAAACAATCTTCGCGTGTGGCACTTCGCGTGAATGAGAGCGAATTCGCGCTTTCGT	780
QY	1288	GACCCGCAAGAGCGCTGGAATCGGGGAGATTATGGAACGTTGGGGGCAACGTCGTCAAG	1347
Db	781	GACCCGCAAGAGCGCTGGAATCGGGGAGATTATGGAACGTTGGGGGCAACGTCGTCAAG	840
QY	1348	CTTTCTGAGTGGCGGTGTGGGCAATTAGGCCCGGCTGTGGCGCTTTGGCGCGCGCATC	1407
Db	841	CTTTCTGAGTGGCGGTGTGGGCAATTAGGCCCGGCTGTGGCGCTTTGGCGCGCGCATC	900
QY	1408	GGGGGCCGGCCCGTGTGCTGATTTGGGCAAGATTCGCGCTTCAACGCTTGGGCGCAGAG	1467
Db	901	GGGGGCCGGCCCGTGTGCTGATTTGGGCAAGATTCGCGCTTCAACGCTTGGGCGCAGAG	960
QY	1468	CTGGGTTTTGCGGCTGTGGCATTTTCCGTGGCGCGGCACTTAAACTTGGCGATGTGTCC	1527
Db	961	CTGGGTTTTGCGGCTGTGGCATTTTCCGTGGCGCGGCACTTAAACTTGGCGATGTGTCC	1020

FEATURES	Db	1528	ATCATGCACACCTTCGCGCGCCGCAATGTGTGCGACGGCGCTGAGAGCTCGGCATCCCAAGC	1587	source	1.300750	/organism="Corynebacterium efficiens YS-314"
	Db	1021	ATCATGCACACCTTCGCGCGCCGCAATGTGTGCGACGGCGCTGAGAGCTCGGCATCCCAAGC	1080		/mol_type="Genomic DNA"	/strain="YS-314"
	Qy	1588	TTCGATTGCGGCGACCTTGTCCAGCTTATGACGCTCCCTCCGCCACGCTTCGGTCAATT	1647		/db_xref="taxon:196164"	/note="CE0885"
	Db	1081	TTCGATTGCGGCGACCTTGTCCAGCTTATGACGCTCCCTCCGCCACGCTTCGGTCAATT	1140	CDS	252..620	/codon_start=1
	Qy	1648	ATTGATGAGGGCGGTTGGCGGTGCGCGCTGCGCCATCTGCGCGCATCTGTCTACGCG	1707		/transl_table=11	/product="hypothetical protein"
	Db	1141	ATTGATGAGGGCGGTTGGCGGTGCGCGCTGCGCCATCTGCGCGCATCTGTCTACGCG	1200		/protein_id="BAC17695.1"	/db_xref="GI:23492723"
	Qy	1708	GCCGAAAAACCGTGTGCTGTCCGCAATTGCCACCGAGGGGGCGCTCGCCATCTCTTCGCG	1767	gene		/translation="WDAQPSGLVDCHYIEFYDTPVPMILFPVGTSGNKRIVLDFNHYFLAKPRILITELPMSCGCLFGWVIGLFTPMHTHYVVCVFCGERDPDSGRITLACLTHASRTMKPCDQGLVANG"
	Db	1201	GCCGAAAAACCGTGTGCTGTCCGCAATTGCCACCGAGGGGGCGCTCGCCATCTCTTCGCG	1260		/gene="rna_168"	/gene="rna_168"
	Qy	1768	GACACCAACACGCGCGGGAATCATGAGCGACGAGCGTGACGCGCAGCATCTTTA	1827	rRNA	529..1936	/gene="rna_168"
	Db	1261	GACACCAACACGCGCGGGAATCATGAGCGACGAGCGTGACGCGCAGCATCTTTA	1320		/gene="rna_168"	/product="16S ribosomal RNA"
	Qy	1828	AGCCAGGGGCTTATCGACGGGATCGTCGCCGAACCGAGCATCTTTGTAGAGAAATTCTC	1887	CDS	complement(2042..2215)	/note="CE0886"
	Db	1321	AGCCAGGGGCTTATCGACGGGATCGTCGCCGAACCGAGCATCTTTGTAGAGAAATTCTC	1380		/codon_start=1	/transl_table=11
	Qy	1888	GGCAATTCAGCAACGCCCTCTCCGAATTGGATTAACAATCCGAGAGGGCGGAGCGGCAC	1947		/product="hypothetical protein"	/protein_id="BAC17696.1"
	Db	1381	GGCAATTCAGCAACGCCCTCTCCGAATTGGATTAACAATCCGAGAGGGCGGAGCGGCAC	1440		/db_xref="GI:23492724"	/translation="MSQTPNSMPPYTHHILGDFADPWWIHPITTCWMPHFGSNTNPPHSHSPMWY"
	Qy	1948	AGTCGCTTCAACGATTGAGCGTTTACGCGCAG	1980	gene	2399..5481	/gene="rna_238"
	Db	1441	AGTCGCTTCAACGATTGAGCGTTTACGCGCAG	1473	rRNA	2399..5481	/gene="rna_238"
RESULT 11	AP005217/c	300750 bp	DNA	linear	BCT 20-DEC-2002		
LOCUS	AP005217	300750 bp	DNA	linear	BCT 20-DEC-2002		
DEFINITION	Corynebacterium efficiens YS-314 DNA, complete genome, section 4/11.						
ACCESSION	AP005217	BA000035					
VERSION	AP005217.1	GI:23492722					
KEYWORDS	Corynebacterium efficiens YS-314						
ORGANISM	Corynebacterium efficiens YS-314						
REFERENCE	1						
AUTHORS	Kawarayashi, Y., Yamazaki, J., Hino, Y., Kikuchi, H., Nakamura, Y., Ikeo, K., Suzuki, M., Mashima, J., Itoh, T., Yamagishi, A., Nishio, Y., Ueda, Y., and Sugimoto, S.						
TITLE	The entire genomic sequence of Corynebacterium efficiens YS-314						
JOURNAL	Published online in Database (2002)						
REFERENCE	2 (bases 1 to 300750)						
AUTHORS	Kawarayashi, Y., Yamazaki, J., Hino, Y., Kikuchi, H., and Director-General of Biotechnology Center.						
TITLE	Direct Submission						
JOURNAL	Submitted (17-MAY-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; Nishihara 2-49-10, Shibuya-ku, Tokyo 151-0066, Japan						
REFERENCE	(E-mail:bioinfr.go.jp, Tel:81-3-3481-1933, Fax:81-3-3481-8424)						
AUTHORS	Kawarayashi, Y. is officially affiliated with the National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, 305-8566 Japan						
TITLE	Nakamura, Y., Ikeo, K., Suzuki, M., and Mashima, J. are at the National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan						
JOURNAL	Itoh, T. is at the Japan Biological Information Research Center, Koto-ku, Tokyo, 135-0064 Japan						
REFERENCE	Yamagishi, A. is at Tokyo University of Pharmacy and Life Science, Hachioji, Tokyo, 192-0392 Japan						
AUTHORS	Nishio, Y., Ueda, Y., and Sugimoto, S. are at the Ajinomoto Co., Inc., Kawasaki, Kanagawa, 210-8661 Japan						
TITLE	The other authors are at the National Institute of Technology and Evaluation, Shibuya-ku, Tokyo, 151-0066 Japan.						
COMMENT	Location/Qualifiers						

TKTSMIIIVNSPHNPTGSVSRSLSEALAGIARAYDLVLSDVEYEHLLTFDKVSHYAV  
ASLPGMWERTITVSSSAKFTENYMTGKALAPALIEAVIRAKQFMSYVGA1PPOPA  
VAHVAHEKPVKORRKGLOKNDILITLTOAGLKVHSHGTFVPAVADIGEDGAEF  
CPDLIDRVGVVAIPVOAFVDPKPEQMSKTRFAFCQEDTLREARQRLRAAGSL"  
complement(7180..7821)  
/note="CE0890"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="BAC17700.1"  
/db\_xref="GI:23492728"

translation="MTESFDVDEAVARRKNEFLDTRKQLQISAEFFGLVLLAIGVL  
YLLADAGVGMWVLLVMGIMALSFMVLPVLEPROVGSASLNTYPLVAVVAEINPR  
DVALLVNTNDELTPRMGLAVRTINLGVDRKLGERSVAVSGRRTFDDQHW  
DEISPMPTMGTDADVVARAKRTIPHDMLNLEKRNRLDEVNTKRMVLLV"  
complement(7822..9495)  
/note="CE0891, similar to A583924-93|CAC31112.1| percent  
identity: 75 in 547 aa"  
/codon\_start=1  
/transl\_table=11  
/product="putative DNA helicase"  
/protein\_id="BAC17701.1"  
/db\_xref="GI:23492729"

translation="MAFGDGLIVQSDKTVLLETGHDLAGEARALAPAELEAREH  
VHTYRTPLALNNARAAGHDAEQVVDLERYSPVPOPLVDVLETMSYGRVVMHK  
HPAGLIESSEPALIVEISRRKIKPMVGQOLDEPPTLAHVSERGRKQKELLTKGMP  
AEDLAGYDEAHPIALSTVEDMSLIRPQOAAAPFEGSGVYVLLPGKARNTVGA  
ASMAQAQTTILVNTYVAGRWKMDLARTLTDEIGERSGEEKLRPVATYQV  
VTRTGERKALELFDSDWGLIYDEVHLPAVFRMTSDQSRRLRGLATVAREG  
GREGVFSLIGPKRYDAPMKDLAQGFATQDCVEVTRTMDAEVMAVATETARV  
LAATHTKTRVVRILIEHQGQPTLIGAYDLQLELAEAPALVDGTSTKKGEL  
FDRFSGELNVLVSKVNFESIDPEAAVAIVGSGTFSRQEARLRLRLRKADGG  
EAHPYSVSRDLTDEYAAHRORFLAEGYARIRIIDADDLEPITEKET"  
complement(9548..11851)  
/note="CE0892, similar to A5006976-11|AAK45126.1| percent  
identity: 36 in 741 aa"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAC17702.1"  
/db\_xref="GI:23492730"

translation="MPCFFSQIRQTKYTSAPKGLDKMKSVPVTLTGMLQDLD  
QLSGILRNRPDLALPPLGSLARLQLRASIQRAVHLSLELAVLEAANAAGAL  
NPITAPVEVERLHRLARSAAYPAETQGEALNRLROLALVGERFTILPEPTMA  
TNMOLLPEHGRSLTRGASOSIEALPARHRIITVLTNSGLGTRPAASADARP  
IPQIKAGLSRVDEQYKLPQYVARVIEGQEDTMRVVPVEPSAVPGADGLAAGL  
EVARLQRMILALSTAPASTLEKALGVAMVRLRLRELQVDEQVAVASLGSGL  
RRGVDDPLPADDDGDYIAPVADENMLQATLQALHMEGWTQTVAAMLVSEADL  
NKPTHLSPASRRDALPETRAMILOSILRVGEEDLEADLPFHHTLAASRIIPERTIRH  
LLEBARWIGAVAGGVTSPARVLTQAPAGVIBETADIAVAVAPKPVNDYFIQADYV  
MVPGLQPEMOKITIGQLADLESFGIASEASLRIAMDLGLTATEIEPTLHSHA  
TGLPOSVAYLLQDLARKHGTLRGGPALCYLRSDDPALHSAVEAAGALGLQIATVA  
AVNSALIEVITALKRAGQPVAAEDLAGSLDAPRAPVPAEOPRSAGPDESTRRA  
AVAAIIRREVARTGTVDQPTLAVLQAAVRGQRTVTLGFVQKGVAVHRTKPLTVNA  
GOVDALDESTGAVHRFMLRITETVVD"  
11836..12036  
/note="CE0893, similar to A583924-91|CAC31110.1| percent  
identity: 50 in 62 aa"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAC17703.1"  
/db\_xref="GI:23492731"

translation="WGNMANVEKGYNDPAMPDHPDSEGHVVTETLIAPYAGASFPWGD  
DMEPVPVPEITIGYVHPYTRINR"  
complement(12229..12822)  
/note="CE0894, similar to AX133781-1|CAC38957.1| percent  
identity: 71 in 195 aa"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAC17704.1"  
/db\_xref="GI:23492732"

translation="MLFMGRHSKATPFPVKAAATTVAFGAANAFAAPASAPSDW  
DLRIACSEGMNIAITNGYGGGLQFSSSTPAVAGDDEPAVPAVATREOOIYAVERT  
LAGQGMWAPACSSSLGINSAPTPPININPAPAPAPAPAPATVDTNTNPVVGSS  
DLNTVOGIYDAVSGLTAQVQVPAETQAHNAIFGR"  
13342..13725  
/note="CE0895, similar to A5006977-4|AAK45135.1| percent  
identity: 52 in 127 aa"  
/codon\_start=1  
/transl\_table=11  
/product="putative cold shock protein"  
/protein\_id="BAC17705.1"  
/db\_xref="GI:23492733"

translation="MPVETVKWYDPERGFGFVSNPGBDFGVKOVLPKGVTEHOGQ  
RIEPPFARGROPALRVKVELPERRRQHTYKPEBLNGIISDWMTLEGTVPALLK  
GHYPRKGEVQAKILRVAAKELEA"  
complement(13861..14408)

Query Match 32.6%; Score 691.4; DB 1; Length 300750;  
Best Local Similarity 65.4%; Pred. No. 146;  
Matches 1046; Conservative 0; Mismatches 546; Indels 7; Gaps 2;

QY	528	TATGCTGGGGGATGAAACACTTCAGCATTTGACGCTCATAGACTCGGTTTGGACCC	587
DB	26058	TAGTGTGAGACATGACACGAACTCCGCCAGGATCTGATCGATTCAGTATGGACAA	25999
QY	588	TGACGCTTCATTTCTTGGAAATGAACTCCCAATATGACAACTCAATCAAGCTATGC	647
DB	25998	GACACGTTTCGCTTCTGGACACTCCCGGAGTACGGAGATGATGATGACTACCG	25939
QY	648	AGAGACTTGGAGCGGGCTCCAAAGCAAGCCAAATGCGATGAATGATTAATCTGAGA	707
DB	25938	GGAGCGCTTGGCCCGCCCGCGAGAAATCCGGGTGATGAGCGGTGATACGGGTGA	25879
QY	708	AGGACCGTGAAGGCAATTCGGTAGCCGTTATTTGTCGATTTTCTTCTCGGCGG	767
DB	25878	GGGAAAGGTGAGGGCTCCCGGTGCGTTCATGCTCCGAGTTGCGCTTCTCGGTGG	25819
QY	768	TTCTTTGGGACAGCGTGGGTGGTGGCATATGAAAGCCATTCACCGCGCACAGACT	827
DB	25818	TTTCATCGTCCGCGACCTCCGCTGCTTATTCACAGGCGCATCCACCGGCGCACGGGA	25759
QY	828	GAACCTCCACTGCTGCTCTCCCTGCTTCCGGTGGTCCGGCATGACAGAAACATCG	887
DB	25758	AAGTCTCCCGCTTTGATCTCGCGGTGTCAGCGGATGACAGAGGGACATCCC	25639
QY	888	AGCTTTGTCATGATGTGTCTCATTAACCGGCGCTGTGACGCGTCAACCGAGCGCATTT	947
DB	25698	GACGTTCCGATGATGATCTCATCAACACCGCGGTCTACCGGACAAAGACGCCACT	25639
QY	948	GCCGTTCTGTGATTTTGGCAATCCAGCATGGGTGGCCCATGCGCTGTGGGGTTC	1007
DB	25638	GCCGTTCTGTGATCTACCTGCTTAACCCCATGAGGTGGGTGATGGCTGTGGGGTTC	25579
QY	1008	ATCTGGGATCTCATCTTTTGGGAACCGGCGCGAGATAGATTTCGGGTCCCGGCT	1067
DB	25578	CGCCGGCATCTTACCTATGCCAGCGGATGACCTGTGGGTTTCTCGGTTCCCGGCT	25519
QY	1068	GCTGAGTTAACCATGCGGATGCGCTTCCAGACGTTGTGACGAGCGGAGAAATTTGCT	1127
DB	25518	GCTGAGTTGACACCGGACCCCATGCGGAGGGGTGACGCGGGCGAGAACCTGCG	25459
QY	1128	GAATACTGTGTGATTTGGAATTTGTGCGCACTCCAAATTTGCTGACGCGGTGCAAA	1187
DB	25458	CGGACCGGTGTCTACACCGGGGTGCTCCCAACCGAGCTGTGGGTGGGTGTCAA	25399
QY	1188	AACCTCAAGGTTATTCAGCGGTAGAGGCAAGGATGTTTCTCCAAACAATCCCTGG	1247
DB	25398	GATCATAGGATTTGGGGAGGGGTGCGCAGGACGCCCCGTAGCTCCGCGACCCGG	25339
QY	1248	CGTGGACCTTCGGTGTATGAGAGGCGATTCGCTTCTGTAACCCGAGAGCTTGAAAT	1307
DB	25338	GAACCAAGTGTGCGTGGGAGTGTGATCATGATCAACCCGGAGAACCGGTTGCGCCGGGGT	25279



```

Qy 1308 CGGGAGATATAGAAAGCTTGGGGGAGACGTCTCAAGCTTTTGTGGCGGTGCTGG 1367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25278 GGGCGATGTGCTGTGCTGCTGTGGGAGAAACGTGTGACCTGTCCGGACAGGTGAGG 25219
Qy 1368 CGCATTTGAGCCCGGTGTGGCGGTTGCCCTGTGGCGCATTCGGAGGCGCGCCGTGTGCT 1427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25218 TCGCATCTCGGGGAGCAACAGGTGGCTGGCCCGCATTCGGGGGCGCTCCGGTGTGCT 25159
Qy 1428 GATTGGGGAGAGA---TCCCGCTTCAAGCTTGGGGCGGAGAGAGTGGCTTTTGGCGGTG 1484
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25158 CATCGGAGAGACCGTCACACCTCCCTCGGGCGCGCGCGCATCGCTTTGCGCCGCG 25099
Qy 1485 TGGCATTTTGTGTCGCGCGGAGACTTAACCTGCGCATCTGTCTCATCATGCAACCTCCGG 1544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25098 CGGATTCGCTCTGGCGGAGAGTGTGAGTGTGCGATCATCAGATCATGACACCCCGG 25039
Qy 1545 CGCCGATTTGTGTCGAGCGGCTGAGAGCTCGGCATGCGCAAGCTGATTCGCGGACCTT 1604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 25038 CGCGAGCTGTGCGGGAGGCGGAGAGAAAGGGCTCGCCGGTTCATTCGCCGACCTTT 24979
Qy 1605 GTCCAAAGTTATCCAGCTCCCTCCGACCGCTTTCGTCATATTGTCAAGGGGTGG 1664
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24978 GGGGAGCTGTGTCAGCGCTCGGTGCCACGCTTCGTCATCTTGAGCGAGGCTGTGG 24919
Qy 1665 CGGTGGCGCGTGGCATGCTGCGCGCGATCTGTCTACGCGGCGGAAAGCGCTGAGCT 1724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24918 AGGCGGTGCTGTGCGCATGTCTCCCGCGACCGGGTGTGGGCGGAGAAAGCTGTGCT 24859
Qy 1725 GTTCGCGATTGCCACGAGAGGCGCTCGGCATCTCTTTCGCGACACCAACACGCGCC 1784
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24858 GTCCCGCTGCGCCCGGAGGGGCGCTCCGCCATCTATACCGTGAACACTCCCAAGCCCC 24799
Qy 1785 GGAATCATAGAGGAGCAAGGCGTGAGGCGGACGACATTTTAAAGCCAAAGGCTTATGA 1844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24798 GAGATGATGAGAACCGGACAGGGGTGTCCGCCATCGATCGTGGCTCCGGGATATGA 24739
Qy 1845 CGGATCTGTGCGGAAACCGACACTTTTGTGAAGAAATTCGCGACAAATCAGCAAGC 1904
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24738 CGGATCTATCGGGAACGATGATTCATGGGGAGGTTGTGGCGGAAATTGACAAAC 24679
Qy 1905 CCTCTCCGAATTGATTAACAATCCGAGAGGGCGGAGCGGACAGTCCCTTACAGATT 1964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24678 CCTCACAATTTGGAAACCAATCCGAGAGGGGTGTGGTGGGAAACAGCCCTTCAGCACTA 24619
Qy 1965 TGAAGCTTAGCGGAGTAAGAAATTAATGCGCTGATCAATGATGATGAACAACAAGG 2024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24618 TGAAGCGCTGGC---TGAATGAACTAGGCGCTGATCAGGTGATGATGAAGACAGGG 24563
Qy 2025 TACGCGCAGACAGTGGTGGCGGAAACCTCAGGCGCTTAAGCAAGCTCTGGCGGAATGG 2084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24562 TGGCGCGGAGACGTGGTGTCCATACCTCTCGGGCGGTAGGCAAGCTTCGGGGGGATGG 24503
Qy 2085 TCAAGTGAACGAGTTCGCGGACCTTCATGCTCTGAATTG 2123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24502 TCAGCTGACGAGCTGCCGCAACTTCATGATGCCGGATGC 24464

```

```

RESULT 12
CGG/TC
LOCUS      C.glutamicum glt gene for citrate synthase and ORF. 3013 bp    DNA     linear    BCT 17-FEB-1995
DEFINITION X66112
ACCESSION  X66112
VERSION    X66112.1 GI:505580
KEYWORDS   citrate synthase; glt gene.
SOURCE     Corynebacterium glutamicum
ORGANISM   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Corynebacteriineae; Corynebacteriaceae; Corynebacterium.
REFERENCE  1 (bases 1 to 3013)
            Eikmanns,B.J., Thum-Schmitz,N., Eggeling,L., Luedke,K.U. and
            Salm,H.
TITLE      Nucleotide sequence, expression and transcriptional analysis of the
            Corynebacterium glutamicum gltA gene encoding citrate synthase

```

```

JOURNAL    Microbiol. 140, 1817-1828 (1994)
REFERENCE  2 (bases 1 to 3013)
AUTHORS    Eikmanns,B.
TITLE      Direct Submission
JOURNAL    Submitted (15-MAY-1992) B. Eikmanns, Institut f Biotechnologie 1,
            Forschungszentrum Juelich GmbH, I B T, Postfach 1913, 5170 Juelich,
            FRG
FEATURES
  source
    1..3013
    /organism="Corynebacterium glutamicum"
    /mol_type="genomic DNA"
    /strain="ATCC13032"
    /db_xref="taxon:1718"
    prim_transcript
    717..>2151
    gene
    838..2151
    /gene="glt"
    /gene="glt"
    /gene="glt"
    /EC_number="4.1.3.7"
    /codon_start=1
    /transl_table=11
    /product="citrate (si)-synthase"
    /protein_id="CAA46902.1"
    /db_xref="GI:505581"
    /db_xref="SWISS-PROT:P42457"
    /translation="MFPRDIIVATDNKKAVALHYPGGEFENDIIEASEGNNGVVLGKMLS
    ETLGTFDPGVSTGTSRESKITIYIDDAIGLRGRYDADLAENATFNEVSYLLNGE
    LPPDELRKENDIRHHTLLDDEPKQFVFPDAPMRLTASVNLSTYYDQNLN
    LDEAQLDKATVRLMAKVPMLAAYAHRRKAPVYPDNSINARENPLRMKFGYPTERY
    EIDPIWKLADKLLIADHEONCSTYVAMISAOAMFVSIAAGINALSGPLHGA
    NOAVLEMLDEDIKSNHGDAIEFNKYNKEDGVRLGFGHRYKNDYDPRRAIYKETAH
    EILEHGLDLDLAIKLEIEIALADYFISRKLYPNVDFTYGLIYRAMGPFPTDFTVL
    FAIGRLPGLMFAHREBQGAAGNKINPROVYTONESRKLVRRER"
    2276..2632
    /note="ORF
    putative binding protein"
    /codon_start=1
    /transl_table=11
    /protein_id="CAA46903.1"
    /db_xref="GI:505048"
    /db_xref="SWISS-PROT:P42458"
    /translation="MEKPIELPVPAPEDLVISDIIVGEABARPGEEVHYVGD
    FETGEBSDSMDGQTSOPFLNGLIAGWDEGIPGMKVGRRULTIPPEAAVGEBSGH
    PLSGRVTFIIDLISA"
  CDS
    716 a      818 c      799 g      680 t
  ORIGIN
    Query Match      24.3%; Score 515; DB 1; Length 3013;
    Best Local Similarity 100.0%; Pred. No. 2.3e-106;
    Matches 515; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  Qy 1609 AAGCTTATCGAGCTGCCCTCCCAACCGTTTGGGTATTTATGTCAGGGCGCTTGGGGGT 1668
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Db 3013 AAGCTTATCGAGCTGCCCTCCCAACCGTTTGGGTATTTATGTCAGGGCGCTTGGGGGT 2954
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Qy 1669 GGGCGCTGGGCAATGTCGCGCGCGATCTGTCTACGCGGCGGAAAGCGTGGCTTCC 1728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Db 2953 GGGCGCTGGGCAATGTCGCGCGCGATCTGTCTACGCGGCGGAAAGCGTGGCTTCC 2894
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Qy 1729 GCATTGGCACAGAGGGCGGCTCGGCATCTCTTCGCGGACCAACAACGCGCGGAA 1788
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Db 2893 GCATTGGCACAGAGGGCGGCTCGGCATCTCTTCGCGGACCAACAACGCGCGGAA 2834
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Qy 1789 ATCATAGAGGACAGAGCGGTGCGAGGCGGACGCACTTTAAGCCAAAGGCTTATCGAGGG 1848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Db 2833 ATCATAGAGGACAGAGCGGTGCGAGGCGGACGCACTTTAAGCCAAAGGCTTATCGAGGG 2774
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Qy 1849 ATGTGCGGGAACCGGACCTTTGTTGAAGAAATCTCGGCACAATCAGCAACGCGCTC 1908
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Db 2773 ATGTGCGGGAACCGGACCTTTGTTGAAGAAATCTCGGCACAATCAGCAACGCGCTC 2714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  Qy 1909 TCCGAATTGATTAACATCCGAGAGGGCGGAGCGGACAGTGCCTTCAACGATTTGAG 1968
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```





Query Match

```

/gene="MT0922"
/note="identified by match to protein family Hm"
/codon_start=1
/transl_table=11
/product="OmpA family protein"
/protein_id="AAK45169.1"
/db_xref="GI:13880481"
/translation="MASHKAGIGOTPATTDARTKPYRSGRPMWIGAVVPIILIA
ICGAFERPOSVPTGVLPITPTSTRASALSLSLSSISGTTVILIGDPDEBA
KAMTALINGLAPGVNVDQIHDPVVASLPSSEPEFTASVPTDGLKVEDTV
TLGTAPSEHKDVAKRAATSTWPMKIVNIEVVGQAPGPAPGACDQASAINAV
TSGPIAFVNDGASLIADYEILNRVADKLKACPDARVINGTDMTSGSGLNIPISAQ
RAKIVADYLVARGVAGDHIAIVGLGSVNPISANATPEGRAKRREIVN"
7426.7578
/gene="MT0923"
7426.7578
/gene="MT0923"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45170.1"
/db_xref="GI:13880482"
/translation="MDFVIVQMSCYLLAFLGSSAVAMVVVTLIKRASHDEGAAPSA
AERTAQ"
7575.8105
/gene="MT0924"
7575.8105
/gene="MT0924"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45171.1"
/db_xref="GI:13880483"
/translation="MMEHVHWMLAGLAFLLGNVLTLMVPEVHQVLKKSXVSSA
KSKPTARKPAVSGTKESEPTAKTKATESAAQIIVAGEPAEPVAGEPAARV
PVVPAYPGPGSARAGADSGPGQVLVGRSDTRLVYPTDPTDPTAQVWPDEBS
AARAPFTPMKSTRRT"
complement(8122..9531)
/gene="MT0925"
complement(8122..9531)
/gene="MT0925"
/note="similar to PID:2440113; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="sensor histidine kinase"
/protein_id="AAK45172.1"
/db_xref="GI:13880484"
/translation="MAGCCIPSAESDSCVCSGDGACMNILRIIPAPPSLRTPVVA
TAIGAIPVLIVGVVWVGITNDRKERDRLDEAGRAIFPVREGDEIRSNDDO
ALITVRGNVTKNSDITLPLKDDYADTVRGVRRVTEIGPEBTSVAVATYD
ATVATNHLNRVLICTFAIGAAVFAWMLAFAVRFQIAQTSIPIDGAPRV
EVHGASEAIEIAEMRWGLQRIWNEONTKELASARFAVSSHEIRTPITARTVL
EVLSTLDPDQKEVLNDVLTQSRLEATSEALERLAQGLSTDHVPDITDLID
RAADPAARIYFDLVSIVSPCTIIVGLPAGRLAVDMAIANVHGAGATVYLSAAS
SRAGEVIAIDNDSGVPEGRQVVFERRSRGTSASHSGSLGLAVVAQQLHGTAS
LENSPLGAKRLVLRIPCS"
complement(9473..10183)
/gene="MT0926"
complement(9473..10183)
/gene="MT0926"
/note="similar to GP:2440114; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA-binding response regulator"
/protein_id="AAK45173.1"
/db_xref="GI:13880485"
/translation="MGGMIDGVTSFPRVLVVDSDSVIASLERRGLSGFEVAVAVDGA
14.7%; Score 312.6; DB 1; Length 17910;

```

```

Best Local Similarity 54.0%; Pred. No. 3e-60;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;
QY 572 ACTCGTTTGGACCCCTGACAGCTTCATTCTTGAATGAACCTCCCAATATGACAAC 631
Db 11770 ACGGGGTCTAGACCGGGGATCTTCTGTCAGCTGGATATACGAGCCCTGGCGTCCGG 11711
QY 632 TCATCAAGGCTATGACAGACCTTGGAGCGGGCTCGAACGAAAGCCAAATGCGATGAT 691
Db 11710 TAGCCGACTCTTAAGCCCGGAGCTGCCCGCTCGGGCGGCACCGCGCGGACAAAT 11651
QY 692 CGTAATATTCAGGAAGGACCGTGAAGGCAATCCGTAGCCGATATTTGCGCAT 751
Db 11650 CGGTGACAGACCGGTGAGGAGCGCTATTCCGCGGGGGTGGCCGTGTGACCTGTAGT 11591
QY 752 TTTCCTTCTCGGCGTTCCTTTGGGACCGGTCCGCTGCTGCGCATATGAGCGCATTC 811
Db 11590 TCGACTTCTGGGGGGCTCGATTGGGTGGACACGCGCAACGATACCGCGCGCTCG 11531
QY 812 ACCGCGCAGACAGCTGAACTCCACTGCTGCTTCCTCGTTCGATGTCGCGCA 871
Db 11530 AGCGGGGACCGCGACCGGCTGCGCTACCTGCGCTACCAAGCTCGGAGGACCCGCA 11471
QY 872 TGCAGAAACAATCGAGCTTTGTATGATGTCATTAACCGGGCTGTGACGCTGTC 931
Db 11470 TGCAGAAAGCAGCGTGCCTTCTGCAATGATGAAGATGCTGCGCATCAGCTGC 11411
QY 932 ACCGCGAGCGCATTTGCGCTTCCTGTGTATTTGGCGCATCCACAGATGGTGGCGCA 991
Db 11410 ACAACGAGCGCGCTGCTTACCTGATTTATTTGGCCATTCGACAGGGTGAAGTTT 11351
QY 992 TGGCTTGTGGGTTATCTGGGCAATCTCATTTGGCGAACCCGGCGCGAGATAGTT 1051
Db 11350 TCGGTGTGGGGCTCGCTGGGGCATCTCACCTGCCGCGCGCTGATGGCT 11291
QY 1052 TCCGAGGTCCTCGCGTGTGAGTTAACCATGCGGATGCGTTCCAGACGAGTGCAGC 1111
Db 11290 TTCTGGGACCAAGGTTATGAGTTGCTGATGCGAACCTTCCTCCATCGGAGTCCAAA 11231
QY 1112 AGCGGAGAAATTTGTGAAAACTGTGTGATGATGAAATGTGTGCCACTTCCAAATGG 1171
Db 11230 CCGCGAGAACTTACCGCGGATGGATCATGACGCGCGTGTGACATGAGACCGGCTAC 11171
QY 1172 GTGACGCGGTGGCAAAACCTCAAGTTATACCGGCTAGAGCGCAAGATGTTTTT 1231
Db 11170 GACGATGCTGATCGTGTGACGCTGCTCATGACCTCCGAAACCGCTTCCGCGCAC 11111
QY 1232 CTCCAACACTCTGCGCTGCA---CTTCCGGTATGAGAGGCAATTTGCGCGTTCTCGTG 1288
Db 11110 CGGAGACGCGCGCGCGCTGACCGATGTCGCCACGTSAGACTGCTGTGATCGCGCC 11051
QY 1289 ACCGCGACAGCGCTGAAATCGGGAGATTAATGAAACGTTGGGGGAGACGTCGTCAAC 1348
Db 11050 GCGCGGACCGCGCGGCTGACGAGCACTGTGACACGCGCGCACGACCGGGTGTGT 10991
QY 1349 TTTCGTGTGCGTGTGCGGATTAAGCCGCGCTGTGCGCTGTGCGCTGTGCGCGCATG 1408
Db 10990 TGTGAGAAACGATCAAGCGGAAGCG---GCGACACGCTGTGCGCTGTGCGCGCTTTG 10934
QY 1409 GGGCGCGCGCGGTGTGATTTGGGCG-----GATCGCGCTTTCAGCG 1453
Db 10933 GCGGCGCAACCGAGGTGTCTCGGCAAGAAAGGCGAGTACGCGCGGGGAGACACTG 10874
QY 1454 TTGGGCTCGAGAGCTGCTTTTGGCGGCTGTGCGATTTGCTGCGCGCGAGCTAAAC 1513
Db 10873 TCGGGCCGCTGTGATTAAGCGAAGCCGACGCGGATGCGCTCGCGCGCGAGCTGTG 10814
QY 1514 TCGCGATCTGTTCATCATGACACTCGCGCGCGGAAATTTGTGCGAGGCGCTGAGAGC 1573
Db 10813 TGCCTGTGTGTGATCATTAAGCGCGCGCGGCGCGCTGTGTGCGCGCGGACGCAAG 10754
QY 1574 TCGGATCCCAAGCTGATTTGGCGGACCTTGTCAAGCTTATGACGCTGCCCTCCCA 1633

```

```

Db      10753 GCGGGCTGGCGCCGACATCGCGCATTCCTGCGCCGACGCTCGTACGCTGATATACCCCA 10694
Qy      1634 CCGTTTCGGTCAATTATTTGGTTCAGAGGCGTTGGCGGCGGTGGCGCGCATGCTGCCGCCG 1693
Db      10693 CCGTTCGATCTCTGCTGGCGCAGGCGCGCGCGCGCGCGCTGGGATGTTGCCGCCGCG 10634
Qy      1694 ATCTGCTACCGCGCGCGGCGGAAACGCGTGGCTGTCCGATTTGCCACAGAGGCGCGCTCG 1753
Db      10633 ACCGGGTGCTGGCGCGCATCTCAAGCGCTGGCGCGCTTGTCTCCGAGAGGCGCAGG 10574
Qy      1754 CCAATCTCTTTCGGCGCAGCACCAACGCGCGCGGAAATCATAGAGCAGCAAGCGCTGACG 1813
Db      10573 CGATGCTGTTCGAGACACTGCTCATGCGCGCACTCGCTGCCGCCCAAGGCGATCGGT 10514
Qy      1814 CGCAGCGCACTTTTAAAGCGAGGCGCTATGACCGGATCGCGCGGAAACCGAGCACTTTG 1873
Db      10513 CGGCGGCACTTACTGAGTGGGCGGATTTGTGACACCATGTCGCGAGTACCCCGACGCCG 10454
Qy      1874 TTGAAGA 1880
Db      10453 CAGACGA 10447

RESULT 14
MTCY31/c      37630 bp      DNA      linear      BCT 03-AUG-2001
LOCUS      Mycobacterium tuberculosis H37Rv complete genome; segment 41/162.
DEFINITION      273101 AL23456
ACCESSION      273101.1 GI:3261565
KEYWORDS
SOURCE
ORGANISM      Mycobacterium tuberculosis H37Rv
                Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
                Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
                tuberculosis complex.
REFERENCE      1 (bases 1 to 37630)
AUTHORS      Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
                Harris, D., Gordon, S.V., Eigmeier, K., Gae, S., Barry III, C.E.,
                Tekle, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
                Connor, R., Davies, R., Devlin, K., Felkell, T., Gentles, S.,
                Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
                Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
                Rajadaram, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
                Squares, S., Squires, R., Sultson, J.E., Taylor, K., Whitehead, S. and
                Barrell, B.G.
                Deciphering the biology of Mycobacterium tuberculosis from the
                complete genome sequence
                Nature 393 (6685), 537-544 (1998)
JOURNAL
MEDLINE      98295987
PUBMED      9634230
REFERENCE      2 (bases 1 to 37630)
AUTHORS      Parkhill, J.
TITLE      Direct Submission
                Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
                tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
                Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
                Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
                75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
                On Jun 27, 1998 this sequence version replaced gi:1314009.
NOTES:
                Details of M. tuberculosis sequencing at the Sanger Centre are
                available on the World Wide Web.
                (URL, http://www.sanger.ac.uk/Projects/M.tuberculosis/) CDS have
                been renumbered from the original cosmid submissions but the old
                gene designations are in brackets after the new gene numbers.
                Gene prediction was based on a Hidden Markov Model of TB genes
                implemented in rtparse (Krogh) supplemented with visual inspection
                of positional base preference in codons, especially where there is
                an increase in the observed/expected third position G + C.
                CAUTION: In some cases we may not have predicted the correct
                initiation codon. Where possible we choose an initiation codon
                (atg, gtg, or ttg) which is preceded by an upstream ribosome
                binding site sequence (optimally 5-13bp before the initiation

```

```

FEATURES
source
codon). If this cannot be identified we choose the most upstream
initiation codon.
Location/Qualifiers
1..37630
/organism="Mycobacterium tuberculosis H37Rv"
/mol_type="genomic DNA"
/strain="H37Rv"
/db_xref="taxon:83332"
/cdate="y31"
/complement(54..1214)
/gene="Rv0874c"
/complement(54..1214)
/gene="Rv0874c"
/complement(54..1214)
/feature="Rv0874c"
/note="Rv0874c. (MTCY31.02c), len: 386, unknown, very
similar to Mycobacterium tuberculosis MTCY20H10_9 (383 aa,
E(): 0, 81.5% identity in 383 aa overlap) and to
hypothetical proteins from SPU62616.1 Synchococcus sp.
(280 aa; E(): 6.3e-26; 35.2% identity in 264 aa overlap)
and SYCSLH102 Synchocystis sp (447 aa; E(): 1.1e-18,
29.5% identity in 400aa overlap)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0874c"
/protein_id="CA97403.1"
/db_xref="GI:1314011"
/db_xref="SWISS-PROT:O10536"
/translation="MRIVGVCTTPDANQAIVAAAGQKRDLAGAPSLAVLGSRAH
TDRADVLSAVLQWIDPVALVGTIAQALVAAHETEDPAAVWVLASGLAETFLDF
VRGSGALITGVRPRTARDLHLDPDYTPPSNLIIEHPYTDLPETAVGVVSGGR
RGGDRLFRDIDVLTSGVGRVLPQMRGVVPSQGRIPGYIVYVAGDGLITELGG
RPLQRLRIEIVGSPDERALVSHGLQIYIVDEHIAAGGDFIIRGLDAPSTGS
IIEIVQVGAITMOQVNDAGADMDLPIETBAAARLPQBAAGALLPTCNGRGRMF
GVABHDASTIELLGGITPLAGFPAAIGTIPAGRNALHGFASMLFPVDME"
/complement(1314..1802)
/gene="Rv0875c"
/complement(1314..1802)
/gene="Rv0875c"
/note="Rv0875c. (MTCY31.03c), len: 162, unknown, possibly
exported protein, highly similar to MLCB57.11 Mycobacterium
leprae cosmid B57, 033056 HYPOTHETICAL 16.1 KD PROTEIN
(162aa). FASTA scores: opt: 789 z-score: 963.4 E(): 0;
71.4% identity in 161 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0875c"
/protein_id="CA97382.1"
/db_xref="GI:1314012"
/db_xref="SWISS-PROT:O10537"
/translation="MKRGVATLPVILVILSVAGAGAWLVRGHPQOPRISYSHG
HLTRVGYLNCNVVDLDDCCPTQAGLPPVSEIRPVQLSVEVISAPEMRLIQTQDP
ANTSTLFRPDRLAVITIPVDPQGRGLTIVQILTLVDSGLRDPVPAHWSVRL
IF"
/complement(1799..3445)
/gene="Rv0876c"
/complement(1799..3445)
/gene="Rv0876c"
/note="Rv0876c. (MTCY31.04c), len: 548, unknown, possible
membrane protein, highly similar to MLCB57.12
Mycobacterium leprae cosmid B57 033057 HYPOTHETICAL 61.7 KD
PROTEIN (579aa). FASTA scores: opt: 2850 z-score: 2863.4
E(): 0; 81.0% identity in 568 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0876c"
/protein_id="CA97383.1"
/db_xref="GI:1314013"
/db_xref="SWISS-PROT:O10564"
/translation="MAPTPGRRTTPGNSVNGHPGMANYPPDDANRYRRRRPPMSNR
YLPVGGQPERPERSVPRTTRAGERITVTRAAAMRSREMSRYTLVHRAATADGAD
KSGITLALTPVPMANFVADSAAVALATLFFAAAGSKSRVALYLLITAPFVAVIP
LIGPALDRLHGRVALDSFGATLALAVLIMNDGATSGFSPVWVYPCALAMVVS
KSFVLSAVTPRMVPTIDLVNRSNLTFTGLIGTIGAGALAAVEFVCTHLEQIP
GALFVVAITTAGSLSMRIPRWVTVSGEVPATLSVHRGRRLRRRMEVKVLCGT

```

LEQPLGRNITTSLMGNCITKRWVGLFLYRPAVAKAHANWVQGLMGLIGAAANG  
 NFGNFTSRRLQDPRPAVLVRCVLTIVLAIIAAVGSLAATLITRAGSALIK  
 ASLDLQSDTLDEESRASGFRSESTLQLANVLGAAGVLYTTELWGFTRVSLALL  
 GLAQIVSRFGSDLLPGLGAGNPPVAAEBOETTRRGAAPQ"  
 complement (3457, .3365)  
 RBS  
 /note="gaggtgac, potential rbs upstream of Rv0876c"  
 3572..3575  
 RBS  
 /note="gag, potential rbs upstream of Rv0877"  
 3583..4371  
 gene  
 /gene="Rv0877"  
 3583..4371  
 /gene="Rv0877c"  
 /note="Rv0877, (MTCY31.05), len: 262, unknown,  
 equivalent to MLCB57\_13 Mycobacterium leprae cosmid B57  
 Q3058 HYPOTHETICAL 28.2 KD PROTEIN (269 aa; E(): 0.80.5\$  
 identity in 257aa overlap.)"  
 /codon\_start=1  
 /transl\_table=1  
 /product="hypothetical protein Rv0877"  
 /protein\_id="CA97384.1"  
 /db\_xref="GI:1314015"  
 /db\_xref="SWISS-PROT:Q10539"  
 /translation="MTGPRESAVATVADWDEGLAAYAVSGADHARAAVVFSGPEAV  
 GYLIVSYSDGNAATRPRAHLPGYQGMQVAVVAVSGADHARAAVVFSGPEAV  
 ABDMVPEERVRPRDLSFDLLPAKDDPRILVPTGTASGDQVDETAAELGLGRWV  
 SSMGAQSGQRWHDGDPGSGAMARSTVRCDCGFPLPLGSLGAMFVCGNELSD  
 GHVIVROYCSQKADSDTTPAGGSTPIYRPYDDGVDLIIEKPAES"  
 complement (4380, .5711)  
 gene  
 /gene="PPE"  
 complement (4380, .5711)  
 CDS  
 /gene="PPE"  
 /note="Rv0878c, (MTCY31.06c), len: 443, unknown; similar  
 to Y56 MYCTU P42611 hypothetical 50.6 kd protein in hsp65  
 3, region, (517 aa), PASTA scores: opt: 1044, z-score:  
 954.9, E(): 0, (47.4% identity in 397 aa overlap). Similar  
 also to MTW014\_3; MTC165\_2, MTC198\_24, MTC1307\_23,  
 MTC148\_17, MTW004\_5, MTW004\_3 etc."  
 /codon\_start=1  
 /transl\_table=1  
 /product="PPE"  
 /protein\_id="CA97385.1"  
 /db\_xref="GI:1314016"  
 /db\_xref="SWISS-PROT:Q10540"  
 /translation="MNFVVLPEVNSARIVAGAGPAPMLAAVAMDIAELGMAAAS  
 FELLISGLTAGSGSAMQCPAAAMAAAPYLSTWNTAKARAEAGAAARAAVVEA  
 AALLAHPRLPAGVNAENOLLSTLVSLPFGONLPAITAEASVEQLMADQVAAVYGG  
 ASTVSQLTPEMQQLSYLPVYVTAAPAGAVCPAALAIPLAGENTICVGFLEING  
 NNVSAGNTGDETFEIGNTIGNANLNGTIGNAGFNFNGNDGNTNFGSN  
 ADFPLNIGSGEDSNLGFQNAADDVTGNGSDGNTNGFNSGDLNTISGPIVQVAG  
 SFGFTGTCGSGFNSGSGSGFQNLGSGSGFQNASPTSGFQNACTALTRASSTA  
 DSRPMPIAPSRLLQVWRTTRATTAECIRIVISRVSTGAPQCKVGNKG"  
 complement (5719, .5721)  
 RBS  
 /note="gga (or aagg) potential rbs upstream of Rv0878c"  
 complement (5985, .6264)  
 gene  
 /gene="Rv0879c"  
 complement (5989, .6264)  
 CDS  
 /gene="Rv0879c"  
 /note="Rv0879c, (MTCY31.07c), len: 91 unknown, equivalent  
 to MLCB57\_14 Mycobacterium leprae O3059 HYPOTHETICAL 9.8  
 KD PROTEIN (91 aa; E(): 1.2e-25, 76.9% identity in 91 aa  
 overlap."  
 /codon\_start=1  
 /transl\_table=1  
 /product="hypothetical protein Rv0879c"  
 /protein\_id="CA97386.1"  
 /db\_xref="GI:1314017"  
 /db\_xref="SWISS-PROT:Q10541"  
 /translation="MSVNSOIRBEPPLPVLLSEWPIAVGALAMVAALVAVVPG  
 LASMPPVYAGATLGLGTLTTFVWQGLAARRGARGAQAAGLETYIDPK"  
 6442..6873  
 gene  
 /gene="Rv0880"  
 6442..6873  
 CDS  
 /gene="Rv0880"  
 6442..6873

Query Match	14.7%	Score 312.6;	DB 1;	Length 37630;
Best Local Similarity	54.0%;	Pred. No. 3e-60;		
Matches 717;	Conservative 0;	Mis. matches 589;	Indels 21;	Gaps 3;
QY 572	ACTCGGTTTTGGACCTCGACAGCTTCATTTCTTGGATTGAAACCTCCCAATATAGACACC	631		
DB 35657	ACGGGGGTGATAGACCGGGGATCTTTTCATCAGCTGGATGACGAGCCCTGCGGTGCCG	3559		
QY 632	TCAATCAAGCTATAGACAGACCTTGGAGCGGGCTTCAMAACAAAGCCAAATGCGATGAT	691		
DB 35597	TAGCCGACTCTTATGCGCGGGAGCTGGCCGCCCTGGGGGGCCACCGGGCCGACGAA	3553		
QY 692	CGGTAAATTACTGAGAAAGGACCGGTGAGGGGCAATTCGGTAGCCGTTATTTTGTCCGATT	751		
DB 35537	CGGTGACAGACCGGTGAGGAGACGGTATTCGGGCGGGCGGGTGGCCGTGTGAGT	3547		
QY 752	TTTCTCTTCTCGGGGTCTTTTGGGACCGGTGGGTGGGTGGCCATATGAAGCCATTC	811		
DB 35477	TCGACTTCTTGGGGGTTCATTTGGGGTGGCAGCGGCCAAGCATACCGCGCCCTCG	3541		
QY 812	ACCGCGCACAGAGCTAAACTCCCACTGTGCTGCCCTGCTCCGTTGGTGGCGGCA	871		
DB 35417	AGCGGGGACCGCCGAGCGGTGCCCTACTTGGCGTATCCAAAGTCCGGAGGCCACCCGCA	3535		
QY 872	TGCAGAAAGCAATCGAGCTTTTGTCTCATGTATGTTTCATTAACCGCGGCTGTGACGCGTC	931		
DB 35357	TGCAAGAAAGGACGGTGGCGCTTCTTCGAGATGGGAAGATCGCTGGGCCATCCAGCTGC	3529		
QY 932	ACCGGAGGGGCATTTGCGCTTCTGTGTATTTTGGCAATCCACGATGGGTGGGCCCA	991		
DB 35297	ACAACCAAGCGCGCTCCCTCACTGTGTCTATTTTGGCCATCCGACCAACGAGTGGAGATTT	3523		
QY 992	TGGCTCTGTGGGGTTCAATCTGGGCATCTCACTTTTGGGAACCGGGCGCCAGATAGATT	1051		
DB 35237	TGGGTGTGGGGCTTCCTCGTGGGCATCTCACCGTCGCGAGCGGGGCGCCCTTATTCGGCT	3517		
QY 1052	TCTGTGGTCTTCGCGGTGTGAGATTAAACAATGGGCAATGGGCTTCGACGAGTGTGAGC	1111		
DB 35177	TTCTTGGGACCAAGGGGTCTATGAGTGTCTATATGGCCACCCCTTCCATTCGGGGTCCAA	3511		
QY 1112	AGCGGAGAAATTTGTTAAAACTGGTGTATTTGATGGAATTTGTGCGCACTCCATTTGC	1171		
DB 35117	CCGCGGAGAAATCTTCGCGCGGCATGGGATTCATCAACGCGCTGTTCGACTGAGCCGGCTAC	3505		
QY 1172	GTTGAGGGGTGGCAAAAACCTCAAGGTATTTACCGCGGTAAGGAGCAAGGATCGTTTT	1231		
DB 35057	GACCGATGTGATCTGCTGCTTGAAGGTGCTCATCGACGCTCCGAAACCGCTTCCGGAC	3499		
QY 1232	CTCCAAACAATCTCTGCGGTGACA---CTTCGCGTATGAGAGCGATTTCCGTTCTCGT	1288		
DB 34997	CGCAGAGCGCCCGGCCCTTACCGGATGTGCAAGTGGGACTGGTGTGGCAATCGGGCC	3493		
QY 1289	ACCGGCAAGGCTGTGAATCGGGAGATTATGAAACGTTGGGGGACAGCTGTCAAGC	1348		
DB 34937	GGCGGACCGGCGGGCGGTCAAGGAGCTACTGCGACACGCGCACACGACCGGGTGTGT	3487		
QY 1349	TTTCTGTGTGGCGGTGCTGGCGCATTTAGACCCCGGTGTGGCGGCTTCCCTTGGCGCGATCG	1408		
DB 34877	TGTAGAGAAACGATCAAGGCGAAAGC---GCGACCAACGCTGTGGCGGTGCGCCGCTTGG	3482		
QY 1409	GGGCGCGGCGCGTGTGCTGATTTGGGAG-----GATCGCGCTTTCACGCG	1453		
DB 34820	CGCGCAACCAACGATGTGTCTCTGGCCAGCAAAAGGACGATAGGCGCGCGGGGAAACACTG	3476		
QY 1454	TTGGCGCGCAAGAGCTGCTTTTTCGCGGTGTGGCAATTTGCTGGCGCGGACCTAAAC	1513		

Db 34760 TCGGGCCGCTGCTTACGCGAAGCCGAGCGGGATGCGCTGCCGCCGAGCTGTGCC 34701  
Qy 1514 TCCCATCGTGTTCATCATGACACCTCCGGCGCGGAATTTGTGACGAGGCGTGAGAGAC 1573  
Db 34700 TCCCGCTGTGCTGCTGATTTAGACCGCGGACCCCGTTGTGCGCGGACGCCAAGCAG 34641  
Qy 1574 TCGGCATCGCAAGCTGATTTGCGCGACCTTTGTCCAAGTTATGACGCTCCCTCCCA 1633  
Db 34640 GCGGGCTGCGCGGACGATCGGCGATTTGCTGCGCGAGCTCGTACAGCTGATACCCCA 34581  
Qy 1634 CCGTTTCGCTATTATTTGCTAGGGCGTTGGCGGTGGCGCGCTGCGCATCTGCCCGCG 1693  
Db 34580 CCGTGTTCGATCTGCTGCGCGCGAGCGCGCGCGCGCGCGCTGCGGATTTGCCCGCG 34521  
Qy 1694 ATCTGCTACGCGCGCGAAGACGCGTGTGCTGCTGCGCATTTGCAACGAGGCGCGCTCG 1753  
Db 34520 ACCGGGTGTGCGCGCGACCTCCAGGCTGTGCGCGCTGCGCTGCGCGCGGAGCGAGCG 34461  
Qy 1754 CCATCTCTTTCGCGCACACCAACGCGCGGAATCATAGACGACGAGCGTGACAG 1813  
Db 34460 CGATGCTGTTCGAGACACTGCTCAATGCGCGCGAATCGCTGCGCGCGAGCATCCGT 34401  
Qy 1814 CGCAGCACTTTTAAAGCAAGGCTTATGACGCGATGCTGCGCGAAGCGACCTTTG 1873  
Db 34400 CGCGCGACTACTGAAATGCGGGATTTGACACCATCGTGTGCGGAGTACCCGACGCG 34341  
Qy 1874 TTGAGA 1880  
Db 34340 CAGACGA 34334

RESULT 15  
LOCUS BX248337/c  
DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 4/14.  
ACCESSION BX248337 BX248333  
VERSION BX248337.1 GI:31617663  
KEYWORDS complete genome.  
SOURCE Mycobacterium bovis subsp. bovis AF2122/97  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.  
REFERENCE 1  
AUTHORS Garnier,T., Elglsheimer,K., Camus,J.-C., Medina,N., Mansoor,H., Pryor,M., Dutfoy,S., Grondin,S., Lacroix,C., Monsemppe,C., Slimen,S., Harriet,B., Atkin,R., Doggett,V., Mayes,R., Keating,L., Wheeler,P.R., Parkhill,J., Barrrell,B.G., Cole,S.T., Gordon,S.V. and Hewinson,G.  
TITLE The complete genome sequence of Mycobacterium bovis  
JOURNAL Online Publication  
REMARK PNAS 10.1073/pnas.1130426100 ( Microbiology )  
REFERENCE 2 (baes 1 to 327650)  
AUTHORS Garnier,T.  
TITLE Direct Submission  
JOURNAL Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex 15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the Mycobacterium bovis sequencing teams, TB Research Group, Veterinary Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone, Surrey KT15 3NB, UK, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK, P4 Annotation, Genopole, Institut Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France. Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France  
FEATURES  
source 1..327650  
location/Qualifiers  
1..327650  
/organism="Mycobacterium bovis subsp. bovis AF2122/97"  
/mol\_type="genomic DNA"  
/strain="AF2122/97"  
/db\_xref="taxon:233413"  
698..2170  
gene

## CDS

/locus\_tag="Mb0912"  
698..2170  
/locus\_tag="Mb0912"  
/note="Mb0912", len: 490 aa. Equivalent to Rv0888, len: 490 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 490 aa overlap). Probable exported protein. Equivalent to AK45157.1 from Mycobacterium tuberculosis strain CDC1551 (597 aa) but shorter 17 aa. Contains possible N-terminal signal sequence."

/codon\_start=1  
/transl\_table=1  
/product="PROBABLE EXPORTED PROTEIN"  
/protein\_id="CAD93773.1"

/db\_xref="GI:31617664"

/translation="MDIAKRTGYQVGAALVLCGAAVTTAIGSAAPDPSSSTDS  
VDACSPLDGSASSLAIIPGASVPOVGARQVDSIPDLALAIIDFLAARNGLVPI  
ENRTPVAPQGVSVBEGTVPVRFDCDPDGNMTFAVRERGAIPGPHQIIVTDQR  
TASFLYADPFGVGTDFSVNVSDTSLHVTGAGYLCPFHGHDVATVTVGNTPT  
DTISGDFSMILVYNIAGLPEPLSAIILPEFVYTKELGKRLNAYVYANVEDPAHQFLI  
KSKMPSQTPPEPPTLMPRLGVPSSDGLNTLSERKVRQLDQTYECTSDNCLTKG  
TYSQRRLDGSQTDVYNIHTHTGSGPTTNALQVANTIQNSAGRAVTVGDEYAR  
SDDSALLQFAQVNGLIDAVQVHGFPTTPEPAVPCWNGNCELIDKLFPSGGQVTL  
QAVSYGNEAPTFENSKGEPISDHPAVGVFHYVDNAVR"  
complement(2205..3326)  
/gene="cItA"  
/locus\_tag="Mb0913c"  
complement(2205..3326)  
/gene="cItA"

## CDS

/locus\_tag="Mb0913c"  
complement(2205..3326)  
/gene="cItA"

/note="Mb0913c, cItA, len: 373 aa. Equivalent to Rv0889c, (100.0% identity in 373 aa overlap). Probable cItA (alternate gene name: gItA), citrate synthase 2 (EC 4.1.3.7), highly similar to others e.g.

CAB95899.1|ALJ59988 putative citrate synthase from Streptomyces coelicolor (387 aa); P39119|CISY BACSU citrate synthase II from Bacillus subtilis (366 aa), FASTA scores: opt: 586, E(): 5.8e-30, (33.8% identity in 367 aa overlap); etc. Also similar to Rv0896|MTCY31.24 from Mycobacterium tuberculosis (29.2% identity in 274 aa overlap) and Rv1131. Contains P500480 citrate synthase signature. BELONGS TO THE CITRATE SYNTHASE FAMILY."

/codon\_start=1  
/transl\_table=1  
/product="PROBABLE CITRATE SYNTHASE II CITA"  
/protein\_id="CAD93774.1"  
/db\_xref="GI:31617665"

/translation="MTVPPENFVGLDGVAFTEIAEPKDGALRYRGDIDVYS  
QRYTFGVMLLVDFNGSGSLPPEPPLIHSDVAVDVQAGLAFPIWGIAPLD  
IDDYTAQQLARASVMAISYVAQSARQIYQPAVQRIIDCSTYTAFFMRQGEPP  
RHEIADIAVYVSAHEGMNASTFARVIASTGADVAAISGALGAMGPHGAPAV  
LPMUDEVERAGDARSVVKIGIDRGKLMGCHRYRAEDPRARVLRAAEELGAPRE  
VAVAVQALSELSELRERPRDRAITNVEFMAVAVLDFARVPAVMMPAVFTGRTAGWCA  
HILEQKRLGLTVRPSAIYVGRPSRPSVQGWERYLTAA"

## gene

complement(3413..6061)  
/locus\_tag="Mb0914c"  
complement(3413..6061)  
/locus\_tag="Mb0914c"

## CDS

/note="Mb0914c", len: 882 aa. Equivalent to Rv0890c, (99.9% identity in 882 aa overlap). Probable transcriptional regulatory protein, luxR family, highly similar (but shorter 238 aa in N-terminus) to NP\_302202.1|NC\_002677 possible transcriptional regulator from Mycobacterium leprae (1106 aa). Also highly similar (generally in part) to others e.g. T50568 probable multi-domain regulatory protein from Streptomyces coelicolor (1334 aa); P10957|NARL ECOLI nitrate/nitrite response regulator protein from Escherichia coli (216 aa), FASTA scores: opt: 193, E(): 6e-06, (37.4% identity in 99 aa overlap); etc. Also highly similar to others from Mycobacterium tuberculosis e.g. MTCV02B10\_22, MTVO08\_44,

MTV036\_21 and MTCY31\_24. Contains P800017 ATP/GTP-binding site motif A (P-loop), P800622 Bacterial regulatory proteins, luxR family signature, and probable helix-turn helix motif from aa 836 to 857 (Score 1559, +4.50 SD). BELONGS TO THE LUXR/UMPA FAMILY OF TRANSCRIPTIONAL REGULATORS. "

gene  
CDS

```

/codon_start=1
/transl_table=1
/product="PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN
(Probable LuxR-family)"
/protein_id="CQJ3775.1"
/db_xref="GI:31617666"
/translation="MRALLAQRNLVTLGCGVGKTRLAIQIASSELRGDLCGVDA
ELTGCGVATLARAAGVLPDQGBSTMSLRFRTGNRMMLWLNCHLIDACALVY
ELTGCGEITILATSRPTGIMAGEITWRPVSMSITDEVLDFPARASVOPGFIAN
NAAVGEICRRLDGPIAIEFAARVRMSPLIEADIDCFRLLAGVGAVQROQT
IRASIDWSMALLTETREQILFRILAIPVGGFLAARVAAGVADDPVSVDITLVD
KLIVVADQCGERTYRLLETYRVALLETIGSGEADVAHRSDVYLAISNTAPDN
DHQIRVARAEMTIDNLRAAFMSRPNNGITTAQLASLQPTWFGRAHLEBGLSWNS
IIEBDRFHLANSTVRARALADKAMSTLWATSPVAGTDLIAPAQQLMARVEGP
AALVALTPACGSSGYNAEAAAPFAEATDILARIIDKMTLCQILWYRGVGTICGSP
NALIRAAEGECRDIALDTIGDRFSRHCSTMLSLAQMWAGNLTLELRSRTAEAEAN
DVPTLRVLEQAOVLVACGASAAHAIAGACIAATELGAGVYOGITGAAMTVAAIAG
DYTALAEASDAPARPILRAPODPOYTHQYLMQLALAGDAILAANQPNADVATNGCH
RMVALLTIRARVATARGEPBELARDDBAHALAGALAHLYIQGPNDDMLLAIGAVSSH
SEGVALLGAALALRQOTRQVRFKIMDAQYQASVYALEBAMQDFDQMAEGALSTD
EAIAAQSGRGERRKRVADARGWSLPTETTERDVRLVSEGSLNKDIAKRLFVSPRTQTHL
THVYALGLASGVLPDQGBSTMSLRTETTERDVRLVSEGSLNKDIAKRLFVSPRTQTHL
complement(6065..6920)
/locus_tag="MB0915C"
complement(6065..6920)

```

gene

CDS

```

/locus_tag="Mb0915C"
/locus="Mb0915C", len: 285 aa. Equivalent to Rv0891C,
len: 285 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.6% identity in 285 aa overlap). Possible
transcriptional regulator, highly similar in N-terminus to
NP_302202.1|NC_002677 possible transcriptional regulator
from Mycobacterium leprae (116 aa). Also highly similar
to several Mycobacterium tuberculosis putative
transcriptional regulators e.g. Q1102|MCYO2B10.22
PROBABLE TRANSCRIPTIONAL REGULATOR PROTEIN (1159 aa),
FASTA scores: opt: 702, E(): 8.3e-40, (50.6% identity in
247 aa overlap); MTv036_21; MTv008_44; MTC02B10.23. Also
shows similarity with several adenylate cyclases and
hydrolases from other organisms."
/codon_table=1
/product="POSSIBLE TRANSCRIPTIONAL REGULATOR PROTEIN"
/protein_id="CAD93776.1"
/db_xref="gi:15171657"
/translation="MLRRAVAVNSLPPIIDIDHALRGEDHPPTCAKCVARGISALGS
LDLRHSLRSCAAPPDVGRSEFVPPRRRVLLAAGLDVSLPTGVTLLADVEST
HLMOCEPDMAVLAHLIDHTVSEALITHNGGQVQPKRKEGSGFLAAFTRASDAACALD
LQRLSLFIRLRIGLHTGEVOLRDELVYGVNTLRALRGAGGVASDAAGDVLV
GLPADAAVLVDLGRHPLRGALPRPEWVQGLCHPDIREFPLFRANKSSPILPAQTT
FVGRRAQTS"
7318..8805
/locus_tag="Mb0916"
7318..8805
/locus="Mb0916"
/locus_tag="Mb0916"
/EC_number="1.14.-.-"
/note="Mb0916, -, len: 495 aa. Equivalent to Rv0892, len:
495 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 495 aa overlap). Probable
monoxygenase (EC 1.14.-.-), highly similar to others e.g.
NP_250787.1|NC_002516 probable flavin-binding
monoxygenase from Pseudomonas aeruginosa (491 aa);
CAB59668.1|AL132674 monoxygenase from Streptomyces
coelicolor (519 aa); P12015|CWO ACIS cyclohexanone
monoxygenase from Acinetobacter sp. (542 aa). FASTA
scores: opt: 489, E(): 6.8e-26, (30.3% identity in 492 aa
overlap); etc. Also highly similar to Rv0555C, Rv3854C,
Rv3083, etc from Mycobacterium tuberculosis. Has
```

hydrophobic stretch at N-terminus."

Query Match	14.7%	Score 312.6;	DB 1;	Length 327650;
Best Local Similarity	54.0%	Pred. No. 3.3e-60;		
Matches 717; Conservative	0;	Mismatches 589;	Indels 21;	Gaps 3;

QY	572	ACTGGTATTTGACCTTCAGAGCTTCATTTCTTGGAATGAAACCTCCCAATATGACAAC	631
Db	21614	ACGGGTGTCTAGACCGGGGAACTTTTCTCAAGCTGGATACGAGCCGCTGGCGTCCGG	21555
QY	632	TCAATCAAGCTATGCAAGACCTTGGAGCGGGCTCGAACAAGCCAAATGCGATGAT	691
Db	21554	TAGCCGACTCTCTATGCCCCGGAGCTGGCCGCCCTCGGGCGGCGACACGGCGCGGACGAAT	21499
QY	692	CGGTAAATTACTGAGGAAGGACCGTGGAGGGCAATCCGTAAGCCGTAATTTTGTCCGATT	751
Db	21494	CGGTGACACCGGTGAGGAGACGGTATTTCCGGCGCGGGTGGCCGTGTGTAGT	21433
QY	752	TTTTCCTTCTCGCGGCTTCTTGTGGGCACGGTCCGTCGTGCGCCATCATGAAGGCAATTC	811
Db	21434	TGCACTTCTGTGGGGGCTCGATTTGGGGTGGCAACGGCCGACGATTCACCGCCGCGTCG	21377
QY	812	ACCGCGCACAGACTGAAACTCCACTGCTGTCTCCCTGCTTCCGTGGTGGCGGCA	871
Db	21374	AGCGGGGACCCGCGACCGGCTCCGCTACCTGACGTCACCAACCTCGGGAGGACCCGCA	21311
QY	872	TGCAGGAAGCAATTCAGAGCTTTTGTGATGATGGTGTCCATTAACGGGGCGTGTCAACGTC	931
Db	21314	TGCAGGAAGCAGGTCCGCTTCTCTGCAAGTGTGAAGATCGTGGCCATCACTATGCG	21255
QY	932	ACCGCGAGGCGCATTTGCTCCGTTCTGGTGTATTTGGCGCAATCCACGATGGTGGCGCA	991
Db	21254	ACAACGAGGCGCGCTGCCCTTACCTGGTCTATTTGGGCATCCGACCAAGGTGGAGTTT	21195
QY	992	TGGCTTGTGGGGTCTATCTGGGCACTCACTTTTGGGAACCCGGCGCGCATATGATT	1051
Db	21194	TGCGTGTGGGGCTCGCTGGGGCACTCAACCGTCCCGAGCGGGCGCCCTGATGGGCT	21133
QY	1052	TCCCTGGGTCTCCGCTGTGTGAATTAACAACCTGGGCAATGGCTTCAGACGGGTGCAGC	1111
Db	21134	TTCTGGGACCAAGGCTCTATGATTTGCTCTATGGCAACCCCTTCCATCCGGGTCCAAA	21077
QY	1112	AGCGGAGAAATTTGTGAACACTGTGTGTATGTATGAAATTTGTGCCCACTTCAATTCG	1171
Db	21074	CCGCCGAGATCTACGGCGGCAATGGATCATCGACGGCGGTGTGCATGTGAACCGGCTAC	21015
QY	1172	GTGCAAGGGGTGGCAAAACCTTAAGTTATTCAGCCGGTAGAGCAACGATCGTTTTT	1231
Db	21014	GACCGATGCTGATCGTGTGTACAGTGTCTCATTCGACCCTCCGAAACCGCTTCCGGCAC	20955
QY	1232	CTCCACAACATCTCTGACGTGTGCA---CTTCCGGTATGAGGCGAATGGCGTTCTCCG	1288
Db	20954	CGGAGACGCCCGGCCCTGATACCCGATGTGTCCACCTGGGAACTCGGTGTGGCATCGGCC	20895
QY	1289	ACCGCGAAGGCTTGAATCGGGAGATTATGAAACGTTGGGGCGAGACGTGTCAACG	1348
Db	20894	GGCCCGACCGGCGGGCGTCAAGGACGCTACTGGACAACGGCGCACCGACCGGGTGTGT	20835
QY	1349	TTTCTGTGCGGTGTGTGCGCATTTAGGCCCGGCTGTGTGGCGTTGTCCCTGGCGCGCATG	1408

```
Db 20834 TGTCAAGAACCGATCAAGGCGAAGCG---GCGACACAGCTGTGCGCTGCGCCCGCTTTG 20778
QY 1409 GGGGCGGCGCCGNGGCTGATTTGGGCGA-----GATGGCCGCTTCAGCGC 1453
Db 20777 GCGGCGAACCCACGCGTGTCTCGGCCACAAAGGCACTAGGCGGCGGGGAGACATG 20718
QY 1454 TTGGGCGCGAGAGCTGCGTTTTGCGCGTCTGCGCATTTGCTGCGCGCGAGCTAAAC 1513
Db 20717 TCGGGCCCGCTCGTTACGCGAAGCCGAGCGGGATGGCGCTCGCGCGCGAGCTGTGCC 20658
QY 1514 TCCCGATCGTGTCCATCATGACACCTCCGCGCGCGAATTGTGCAAGCGCGCTGAGAGC 1573
Db 20657 TCCCGCTGTGTGTGATTTGAAGCGCGCGACCCGCGTGTGTGCGCGCAGCGAACAGG 20598
QY 1574 TCGGATGCGAAGCTGATTTGCGCGCACCTTGTCCAAAGCTTATCGACGCTCCCTCCCA 1633
Db 20597 GCGGCGTGGCGCGCACGATCGCGCATTTGCTGCGCGAGCTCGTCAAGCTGATACCCCGA 20538
QY 1634 CCGTTTCGATCATTAATTGCTCAGGGCGTTGCGCGTGGCGCGCTGGCATGTGCCGCG 1693
Db 20537 CCGTGTGATCTGTGCTGGGCGAGGGGAGCGGCGCGCGCTGGGATGTGCCGCG 20478
QY 1694 ATCTGCTCTACGCGGCGGAAACGCGTGTGTCTCGCATTTGCCACCAAGGCGGCTCGG 1753
Db 20477 ACCGGGTGTGCGCGCATCTCAAGCGCTGTGCGCGCTTGTCTCCGAAGGAGCAAGCG 20418
QY 1754 CCATCTCTTCCGCGACACCAACCAAGCGCGGAAATCATAGAGGACAAAGCGTGCAGG 1813
Db 20417 GATGTGTGTGAGAGACATGCTCATGCGCGCAACTGCTGCGCGCCCAAGGCAATCCGT 20358
QY 1814 GCGACGCACTTTTAAGCCAAAGGCTTATGACGCGGATCGTCCGCGAAACCGAGCACTTG 1873
Db 20357 CCGCGGACCTACTGAAGTCGGGGATTTGTGACACCATGTGCGCGAGTACCCCGAGCGCG 20298
QY 1874 TTGAAGA 1880
Db 20297 CAGACGA 20291
```

Search completed: November 13, 2003, 00:10:28  
Job time : 7983.97 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 15:05:16 ; Search time 588.017 Seconds  
(without alignments)  
9746.164 Million cell updates/sec

Title: US-10-024-370-1

Perfect score: 2123

Sequence: 1 ctccgagcgagtcgctgctatc.....gacctcatgcctggaattc 2123

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq.190u03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2123	100.0	2123	22	AAC82732
2	2111	99.4	349980	22	AAH68526
3	1575	74.2	1575	22	AAH67755
4	1473	69.4	1473	22	AAH65894
5	1473	69.4	1473	22	AAC82733
6	312.6	14.7	1488	22	AAH51981
7	312.6	14.7	4403765	22	AAI96683
8	312.6	14.7	4411529	22	AAI96682

9	168	7.9	168	25	ACA00248
10	131	6.2	354	22	AAH68407
11	122.8	5.8	873	23	AAH54163
12	122.2	5.8	876	25	ABZ39000
13	121	5.7	867	23	AAH53191
14	121	5.7	6021	20	AAH13395
15	121	5.7	6021	24	ABH99190
16	119.2	5.6	676	24	ABH65750
17	119.2	5.6	870	23	AAH53651
18	119.2	5.6	877	24	ABH65721
19	119.2	5.6	1053	19	AAH14153
20	118.6	5.6	349980	21	AAH21608
21	118.6	5.6	1437668	21	AAH81490
22	108.6	5.1	23673	23	ABZ75344
23	108	5.1	993	23	AAH56217
24	102.8	4.8	915	23	AAH52496
25	99.8	4.7	3118	23	AAH88487
26	99.8	4.7	3118	23	AAH89710
27	97.8	4.6	867	23	AAH55883
28	96.8	4.6	654	24	ABH78240
29	96.2	4.5	864	25	ABH06060
30	96.2	4.5	867	22	AAH90763
31	96.2	4.5	891	23	AAH53478
32	96.2	4.5	1595	19	AAH296348
33	96.2	4.5	19702	19	AAH52140
34	96.2	4.5	1810121	17	AAH20653
35	96.2	4.5	2162598	25	ABH66454
36	94.2	4.4	873	24	ABH68650
37	94.2	4.4	2155561	24	ABH71527
38	93.6	4.4	495269	24	ABQ67195
39	93.6	4.4	3011208	24	ABH069245
40	93	4.4	867	23	AAH55630
41	93	4.4	891	22	AAH90875
42	92.8	4.4	1001	22	AAH91440
43	92.8	4.4	1001	22	ABH37820
44	91	4.3	2365589	24	ABA90521
45	88	4.1	2944528	24	ABA03041

#### ALIGNMENTS

RESULT 1	
AAC82732	standard: DNA; 2123 BP.
AC	AAC82732;
XX	
DT	15-MAR-2001 (first entry)
XX	
DE	C. glutamicum accDA DNA.
XX	
KW	L-amino acid: accDA; coryneform microorganism; L-lysine; animal feed; medicine; pharmaceutical industry; ds.
XX	
OS	Corynebacterium glutamicum.
XX	
FN	EP1055725-A2.
XX	
PD	29-NOV-2000.
XX	
PF	10-MAY-2000; 2000EP-0109842.
XX	
PR	27-MAY-1999; 99DE-1024365.
XX	
PA	(DEGS) DEGUSA-HUELS AG.
XX	
PA	(KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
XX	
PI	Tilg Y, Eggeling L, Eikmanns B, Sahn H, Moeckel B,
XX	
DR	WPI, 2001-042411/06.
XX	
DR	P-PsDB; AAB45789.
XX	

C. glutamicum deri  
C. glutamicum codin  
Pseudomonas aerugi  
N. gonorrhoeae nuc  
Enterococcus faeca  
Enterococcus faeca  
Enterococcus faeca  
Helicobacter pylori  
Helicobacter pylori  
Helicobacter pylori  
H. pylori GHP0.451  
Neisseria meningit  
N. meningitidis B  
Human R128 gene C  
Salmonella typhi D  
E. coli DNA for ce  
DNA encoding novel  
DNA encoding novel  
Streptococcus pneu  
Bacillus clausii g  
S. pneumoniae type  
CFE 67 coding sequ  
Haemophilus influe  
S. pneumoniae deri  
Streptococcus pneu  
Haemophilus influe  
Streptococcus pneu  
Streptococcus poly  
Listeria innocua c  
Listeria innocua D  
Streptococcus pneu  
2CFE 67 coding seq  
Haemophilus influe  
DNA sequence upstr  
Genomic sequence o  
Listeria monocytog

PT Cloned Corynebacterium glutamicum accDA gene useful for producing  
PT transformed coryneform bacteria producing increased yields of L-amino  
acids, especially L-lysine

PS Claim 1; Page 9-11; 20pp; German.

XX The invention describes a novel cloned Corynebacterium glutamicum DNA  
XX (1), replicable in coryneform microorganisms coding for an accDA gene  
XX The invention also describes (1) a polypeptide derived from (1) having  
XX a fully defined 491 aa sequence; (2) coryneform microorganisms  
XX transformed with one or more copies of (1); (3) the shuttle vector  
XX pZlacADA contained in Corynebacterium glutamicum DSM 12785; and  
XX (4) a process for producing L-amino acids, comprising culturing a  
XX coryneform bacterium that overexpresses the accDA gene. Coryneform  
XX bacteria transformed with (1) so that they overexpress the accDA gene are  
XX useful for producing L-amino acids, especially L-lysine, which are  
XX useful in animal feeds, in human medicine and in the pharmaceutical  
XX industry.

Sequence 2123 BP; 460 A; 574 C; 592 G; 497 T; 0 other;

Query Match 100.0%; Score 2123; DB 22; Length 2123;

Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 CTCGAGCGGAGATGCGGTGATCGGCACTCTTAAGCAATGCCGGTTTAAATAAGCA 60
DB 1 CTCGAGCGGAGATGCGGTGATCGGCACTCTTAAGCAATGCCGGTTTAAATAAGCA 60
QY 61 CTTATATGTTTCTACACACATCTGGCCGACGACGAGATATGTTTGCATCAGCTA 120
DB 61 CTTATATGTTTCTACACACATCTGGCCGACGACGAGATATGTTTGCATCAGCTA 120
QY 121 AACGTGATGATGTAAGTTACTTAATCTCAATTCATTCGATGCGATGCGATGCGA 180
DB 121 AACGTGATGATGTAAGTTACTTAATCTCAATTCGATGCGATGCGATGCGA 180
QY 181 ACTGCCCAATATCTTAACCTTAACTTAAAGTATGTTTACTGATTAATAAGT 240
DB 181 ACTGCCCAATATCTTAACCTTAAAGTATGTTTACTGATTAATAAGT 240
QY 241 TCCCGATCTACCCCTCTTTACCCCGAATAACCCCTTTGCAAGATTGCAACAACA 300
DB 241 TCCCGATCTACCCCTCTTTACCCCGAATAACCCCTTTGCAAGATTGCAACAACA 300
QY 301 GTGCAATGTTAAAGGCTTCAACACGTCACATTCGTCCGCTTTAGGCTATGTCGG 360
DB 301 GTGCAATGTTAAAGGCTTCAACACGTCACATTCGTCCGCTTTAGGCTATGTCGG 360
QY 361 ACGTCTAGGCAAAAAGTATGTTGAGATGAAGAACGCAATCCGTCATTTTAAAGCA 420
DB 361 ACGTCTAGGCAAAAAGTATGTTGAGATGAAGAACGCAATCCGTCATTTTAAAGCA 420
QY 421 TCGATAGCTTAATTTGGGCTTAGATCTTCGCGCTTAAATAGTATGACAGACATTGA 480
DB 421 TCGATAGCTTAATTTGGGCTTAGATCTTCGCGCTTAAATAGTATGACAGACATTGA 480
QY 481 ATTAAATTAACAAGCCATTTTTCGCGCTGAGAAAGCTTTCCGACTATGTCGGGC 540
DB 481 ATTAAATTAACAAGCCATTTTTCGCGCTGAGAAAGCTTTCCGACTATGTCGGGC 540
QY 541 ATGGAACACACTTACGATGAGCTGATAGCTCGGTTTGGACCCGGAAGCTTCAAT 600
DB 541 ATGGAACACACTTACGATGAGCTGATAGCTCGGTTTGGACCCGGAAGCTTCAAT 600
QY 601 TCTTGAATGAACCTCCCAATATGACAACTCAATCAAGGCTATGAGAGACCTTGAG 660
DB 601 TCTTGAATGAACCTCCCAATATGAGAACTCAATCAAGGCTATGAGAGACCTTGAG 660
QY 661 CGGGCTGAGAGCAAGGCAATGAGATGCGTAATTAATGAGAGAGCAGCTGGAG 720
DB 661 CGGGCTGAGAGCAAGGCAATGAGATGCGTAATTAATGAGAGAGCAGCTGGAG 720
```

```
QY 721 GGCAATCCGGTAGCCGTTATTTTGTCCGATTTTCTTCTCGCGGTTCTTTGGGACG 780
DB 721 GGCAATCCGGTAGCCGTTATTTTGTCCGATTTTCTTCTCGCGGTTCTTTGGGACG 780
QY 781 GTGCGTCCGGTGGCAGATCAATGAAGGAGATTCACCGGCGCACAGAGCTGAATCCCACTG 840
DB 781 GTGCGTCCGGTGGCAGATCAATGAAGGAGATTCACCGGCGCACAGAGCTGAATCCCACTG 840
QY 841 CTGGTCTCCCTGTCTTCGATGATGCGCGCATGACAGAAACAATCAGCTTTTGTATG 900
DB 841 CTGGTCTCCCTGTCTTCGATGATGCGCGCATGACAGAAACAATCAGCTTTTGTATG 900
QY 901 ATGATGTCATTAACCGGCGCTGTGACAGCTTACCGGAGCGCATTTGCCGTTCTG 960
DB 901 ATGATGTCATTAACCGGCGCTGTGACAGCTTACCGGAGCGCATTTGCCGTTCTG 960
QY 961 TATTTGGCAATCCCAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
DB 961 TATTTGGCAATCCCAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
QY 1021 ACTTTTGGGAAACCCGCGGCGAGATAGTATGTTTCTGAGGTTCTGCGTGTGAGTTAAC 1080
DB 1021 ACTTTTGGGAAACCCGCGGCGAGATAGTATGTTTCTGAGGTTCTGCGTGTGAGTTAAC 1080
QY 1081 ACTGGGATGCGCTTCCAGCGGTGTGACAGAGGCGGAGAAATTTGGTGAATACTGGTGTG 1140
DB 1081 ACTGGGATGCGCTTCCAGCGGTGTGACAGAGGCGGAGAAATTTGGTGAATACTGGTGTG 1140
QY 1141 ATTGATGGAATTTGTGTGCGCACTCAATTGCGTGCAGCGGATGCAAAAACCTCAAGTT 1200
DB 1141 ATTGATGGAATTTGTGTGCGCACTCAATTGCGTGCAGCGGATGCAAAAACCTCAAGTT 1200
QY 1201 ATTGAGCGGATGAGGCAAGCAATGATGTTTCTCAAAACCTCTGAGCGTGCATTCG 1260
DB 1201 ATTGAGCGGATGAGGCAAGCAATGATGTTTCTCAAAACCTCTGAGCGTGCATTCG 1260
QY 1261 GTGATGAGGCGAATTCGCGCTTCTGTAACCCGCAAGGCTTGGATTCGGGAGATTATG 1320
DB 1261 GTGATGAGGCGAATTCGCGCTTCTGTAACCCGCAAGGCTTGGATTCGGGAGATTATG 1320
QY 1321 GAAACGTTGGGCGACAGCTGCTCAAGCTTTCTGTGCGGCTCTGAGCGATGAGCCG 1380
DB 1321 GAAACGTTGGGCGACAGCTGCTCAAGCTTTCTGTGCGGCTCTGAGCGATGAGCCG 1380
QY 1381 GCTGTGCGCTTCCCTGCGCGCATGCGGAGCGCGCGCTGTGCTGATTTGGGAGAT 1440
DB 1381 GCTGTGCGCTTCCCTGCGCGCATGCGGAGCGCGCGCTGTGCTGATTTGGGAGAT 1440
QY 1441 CGCGGCTTCAAGCTTGGGCGCGAGAGCTGCTTGGCGGCTGCTGCTGCTGCTGCTG 1500
DB 1441 CGCGGCTTCAAGCTTGGGCGCGAGAGCTGCTTGGCGGCTGCTGCTGCTGCTGCTG 1500
QY 1501 CGGAGCTTAACTGCGCATGCTGTCATCATGACACCTCGCGCCGCAATTTGTGCA 1560
DB 1501 CGGAGCTTAACTGCGCATGCTGTCATCATGACACCTCGCGCCGCAATTTGTGCA 1560
QY 1561 GCGGCTGAGAGCTCGGATGCGAATGCGAATGCGGACCTTGTCCAAGCTTATGAC 1620
DB 1561 GCGGCTGAGAGCTCGGATGCGAATGCGAATGCGGACCTTGTCCAAGCTTATGAC 1620
QY 1621 GCTCCCTCCCAACGCTTGGCTTATTTGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
DB 1621 GCTCCCTCCCAACGCTTGGCTTATTTGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
QY 1681 ATGCTGCGCGCATGCTGCTTACGCGCGGAAACGCTGCTGTCCGATTTGCCACCA 1740
DB 1681 ATGCTGCGCGCATGCTGCTTACGCGCGGAAACGCTGCTGTCCGATTTGCCACCA 1740
QY 1741 GAGGCGCTTCCGCTTCTTCCGCAACCAACGCGCGGAAATCAATGAGCGA 1800
DB 1741 GAGGCGCTTCCGCTTCTTCCGCAACCAACGCGCGGAAATCAATGAGCGA 1800
QY 1801 CAAGGCGTGAAGCGGACGACCTTTTAAGCCAAAGGCTTATGAGCGGATTCGTCCGCGA 1860
```

```

Db      1801  CAAAGCGTGCAGCGCAGCAGCTTTAAGCAGAGGGCTTATGACGGATCGTCCGAA 1860
Qy      1861  ACCGACATTTGTTGAAGAAATTCCTCGGACAAATAGCAAGCCCTCCCAATTGAT 1920
Db      1861  ACCGACATTTGTTGAAGAAATTCCTCGGACAAATAGCAAGCCCTCCCAATTGAT 1920
Qy      1921  AACCAATCCGAGAGGGCGGAGCGCAGTCTGCTCAGACGATTTAGGCGAG 1980
Db      1921  AACCAATCCGAGAGGGCGGAGCGCAGTCTGCTCAGACGATTTAGGCGAG 1980
Qy      1981  TAAAGAAATTAATGCGCTGATCAATGATGAACACGAGGTAACGCGACAGTGG 2040
Db      1981  TAAAGAAATTAATGCGCTGATCAATGATGAACACGAGGTACGCGCAGACAGTGG 2040
Qy      2041  GTGGCGGAGACCTCAGAGGGCGCTTAAGCAGCCTCTGGCGGAATGTTAGTACAGTCC 2100
Db      2041  GTGGCGGAGACCTCAGAGGGCGCTTAAGCAGCCTCTGGCGGAATGTTAGTACAGTCC 2100
Qy      2101  GCCGACCTTCATGCTGGAATTC 2123
Db      2101  GCCGACCTTCATGCTGGAATTC 2123

```

RESULT 2  
AAH68526/c  
ID AAH68526 standard; DNA; 349980 BP.

AAH68526;

26-SEP-2001 (first entry)

C glutamicum coding sequence fragment SEQ ID NO: 7061.

Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

organic acid synthesis; ds.

Corynebacterium glutamicum.

EP108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-0127688.

16-DEC-1999; 99JP-0377484.

07-APR-2000; 2000JP-0159162.

03-AUG-2000; 2000JP-0280988.

(KYOWA) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

Tateishi N, Senoh A, Ikeda M, Ozaki A;

WPI; 2001-376931/40.

Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analysing expression profile or pattern of a gene and identifying homologous gene

Disclosure; SEQ ID NO: 7061; 246bp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Coryneform bacterium, and identifying a homologue of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from the CC European Patent Office.

XX SQ Sequence 349980 BP; 79274 A; 90638 C; 98727 G; 81341 T; 0 other;

Query Match 99.4%; Score 2111; DB 22; Length 349980;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2122; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

Qy      1  CTGAGCGGAGTCGGTATCGGCCCATCTCTTAAGCAATGCCGCTTTAAATAAGCAA 60
Db      281620  CTGAGCGGAGTCGGTATCGGCCCATCTCTTAAGCAATGCCGCTTTAAATAAGCAA 281561
Qy      61  CTTATATGTTCTCAACATCTGGCCGACGACGAAATGTTTTCATCAAGCTA 120
Db      281560  CTTATATGTTCTCAACATCTGGCCGACGACGAAATGTTTTCATCAAGCTA 281501
Qy      121  AACGTGTAATGTGAATTAACCTAATCAATTCGAATGCGATTCGATTTGGAATCTC 180
Db      281500  AACGTGTAATGTGAATTAACCTAATCAATTCGAATGCGATTCGATTTGGAATCTC 281441
Qy      181  ACTCCCCCAATATCTTAACCTTAACTTAAAGTAGTGTCTTACCTGATTTATTAAGT 240
Db      281440  ACTCCCCCAATATCTTAACCTTAACTTAAAGTAGTGTCTTACCTGATTTATTAAGT 281381
Qy      241  TCCCGATCTACCCCTCTTTACCCGAAATACCCCTTTTGCAAAATTTGCAAAACAACA 300
Db      281380  TCCCGATCTACCCCTCTTTACCCGAAATACCCCTTTTGCAAAATTTGCAAAACAACA 281321
Qy      301  GTGCAATAGTTAAACGGGCTTCAACAGTCAACATTCGTCGGTTTATAGCTATGTTCCGG 360
Db      281320  GTGCAATAGTTAAACGGGCTTCAACAGTCAACATTCGTCGGTTTATAGCTATGTTCCGG 281261
Qy      361  ACGTCTAGCAAAAAGTAGTGTGTGAGATGAAGACATTAATCCGTCAATTTTACGCA 420
Db      281260  ACGTCTAGGC-AAAAGTAGTGTGTGAGATGAAGACATTAATCCGTCAATTTTACGCA 281202
Qy      421  TCGATAGCTTAATTTGGGCTTAGATCTTCGCGCTTAATATGATGACAGACATTTGA 480
Db      281201  TCGATAGCTTAATTTGGGCTTAGATCTTCGCGCTTAATATGATGACAGACATTTGA 281142
Qy      481  ATTAATTAACAAAGCATTTTTCGCGCGTGAAGAGCGTTTTCGATATGCTGAGGAGC 540
Db      281141  ATTAATTAACAAAGCATTTTTCGCGCGTGAAGAGCGTTTTCGATATGCTGAGGAGC 281082
Qy      541  ATGGAACACACTTCAGATTCAGCTCATAGACTCGTTTGAACCTTGAACAGCTTCATT 600
Db      281081  ATGGAACACACTTCAGATTCAGCTCATAGACTCGTTTGAACCTTGAACAGCTTCATT 281022
Qy      601  TCTTGAATGAATCTCCCAATATGACAACTCAATCAAGCTATGACAGACCTTGGAG 660
Db      281021  TCTTGAATGAATCTCCCAATATGACAACTCAATCAAGCTATGACAGACCTTGGAG 280962
Qy      661  CGGGCTGGAAGAGGCGCAATATGATATCGGTAATTAATGGAAGAGGACCGTGGAG 720
Db      280961  CGGGCTGGAAGAGGCGCAATATGATATCGGTAATTAATGGAAGAGGACCGTGGAG 280902
Qy      721  GGCATTCGGTAGACCGTATTTTGTCCGATTTTCTCTCGCGGCTTCTTTGGGACAG 780
Db      280901  GGCATTCGGTAGACCGTATTTTGTCCGATTTTCTCTCGCGGCTTCTTTGGGACAG 280842
Qy      781  GTCCGCTGCGTGGCGCATATGAAGCGATTACCGCGCAGACAGAGCTGAAACTCCACTG 840
Db      280841  GTCCGCTGCGTGGCGCATATGAAGCGATTACCGCGCAGACAGAGCTGAAACTCCACTG 280782
Qy      841  CTGGTCTCCCGTCTGCTCCGCTGCGTGGCGCATATGGAAGAGCAATTCAGCTTTGTGATG 900
Db      280781  CTGGTCTCCCGTCTGCTCCGCTGCGTGGCGCATATGGAAGAGCAATTCAGCTTTGTGATG 280722
Qy      901  ATGCTGTCATTAACCGGCTGTGCAAGGTGACCGGAGGCGCATTTGCTTCTGCTG 960
Db      280721  ATGCTGTCATTAACCGGCTGTGCAAGGTGACCGGAGGCGCATTTGCTTCTGCTG 280662

```

QY 961 TATTGGCCAAATCCAGAGATGGGTGGCGCCATGGGCTCGTGGGGTTCAATCTGGGCAATCTC 1020  
DB 280661 TATTGGCCAAATCCAGAGATGGGTGGCGCCATGGGCTCGTGGGGTTCAATCTGGGCAATCTC 280602  
QY 1021 ACTTTTGGGGAACCCGGCGCGAGATAGGTTTCTGGGTCTCGCGGTGGAGTTAAC 1080  
DB 280601 ACTTTTGGGGAACCCGGCGCGAGATAGGTTTCTGGGTCTCGCGGTGGAGTTAAC 280542  
QY 1081 ACTGGGCAATGGGCTTCAGACGCTGTGACAGCGGAGAAATTTGTGAAAATCTGTGTG 1140  
DB 280541 ACTGGGCAATGGGCTTCAGACGCTGTGACAGCGGAGAAATTTGTGAAAATCTGTGTG 280482  
QY 1141 ATTGATGAATTTGTGTCGCACTCCAAATTTGCGTGGAGGGTGGGAAAAACCTCAAGTT 1200  
DB 280481 ATTGATGAATTTGTGTCGCACTCCAAATTTGCGTGGAGGGTGGGAAAAACCTCAAGTT 280422  
QY 1201 ATTCAGCCGGTAGAGCAACGAGATGTTTCTCCAAACAATCTGGCGTGGCACTTCCG 1260  
DB 280421 ATTCAGCCGGTAGAGCAACGAGATGTTTCTCCAAACAATCTGGCGTGGCACTTCCG 280362  
QY 1261 GTGATGAGGCGAATTCGCGTTCTGTGACCCGCAAGAGGCTTGAAATCGGGAATTATG 1320  
DB 280361 GTGATGAGGCGAATTCGCGTTCTGTGACCCGCAAGAGGCTTGAAATCGGGAATTATG 280302  
QY 1321 GAAACGTTGGGGGAGAGGTGTCAGAGCTTCTGGTGGCGCGTGGGCGCAATTGAGCCG 1380  
DB 280301 GAAACGTTGGGGGAGAGGTGTCAGAGCTTCTGGTGGCGCGTGGGCGCAATTGAGCCG 280242  
QY 1381 GCTGTGCGCGTGTGCCCTGTGCGCGCATCGGAGGCGCGGCCCGTGTGCTGATTTGGGAGGAT 1440  
DB 280241 GCTGTGCGCGTGTGCCCTGTGCGCGCATCGGAGGCGCGGCCCGTGTGCTGATTTGGGAGGAT 280182  
QY 1441 CGCCGCTTCAGGCTTTGGGCGCAGAGAGCTGGGCTTTGGCGCTGTGGCAATTTCCGTGGCG 1500  
DB 280181 CGCCGCTTCAGGCTTTGGGCGCAGAGAGCTGGGCTTTGGCGCTGTGGCAATTTCCGTGGCG 280122  
QY 1501 CGCGAGCTAAACCTGCGCGATGTGTCATTCATTCGACACTCCGGGCGGAAATGTGCGAG 1560  
DB 280121 CGCGAGCTAAACCTGCGCGATGTGTCATTCATTCGACACTCCGGGCGGAAATGTGCGAG 280062  
QY 1561 GCGGCTGAGAGCTCGGAGATCGAATTCGCGCACTTGTCCAGCTTAATCGAC 1620  
DB 280061 GCGGCTGAGAGCTCGGAGATCGAATTCGCGCACTTGTCCAGCTTAATCGAC 280002  
QY 1621 GCTCCCTCCCAACCTGTTCCGTCAATTAATGTCAGGCGCTGGCGCGCTGGCC 1680  
DB 280001 GCTCCCTCCCAACCTGTTCCGTCAATTAATGTCAGGCGCTGGCGCGCTGGCC 279942  
QY 1681 ATGCTGCCCGCGATCTGTCTAAGCGGCGGAAACGCGTGGCTGTCCGCAATTCGACCA 1740  
DB 279941 ATGCTGCCCGCGATCTGTCTAAGCGGCGGAAACGCGTGGCTGTCCGCAATTCGACCA 279882  
QY 1741 GAGGCGCGCTCGGCGCATCTTCCGCGACAACAACAACGCGCGGAAATATAGAGCA 1800  
DB 279881 GAGGCGCGCTCGGCGCATCTTCCGCGACAACAACAACGCGCGGAAATATAGAGCA 279822  
QY 1801 CAAGGCGTGCAGGCGCGACACTTTTAAGCAAGGGCTTAATGACGGGATGTCGCCGAA 1860  
DB 279821 CAAGGCGTGCAGGCGCGACACTTTTAAGCAAGGGCTTAATGACGGGATGTCGCCGAA 279762  
QY 1861 ACCGAGCACTTTGTTGAAGAAATCTGGGCAATTCAGCAACGCGCTCCGAAATTTGAT 1920  
DB 279761 ACCGAGCACTTTGTTGAAGAAATCTGGGCAATTCAGCAACGCGCTCCGAAATTTGAT 279702  
QY 1921 AACCAATCCGAGAGGCGCGAGCGCAAGTGCCTTCACACGATTGAGCGTTAGCGGAG 1980  
DB 279701 AACCAATCCGAGAGGCGCGAGCGCAAGTGCCTTCACACGATTGAGCGTTAGCGGAG 279642  
QY 1981 TAAAGAAATTAATGCGCTGATCAATTCGATATGAAACAACGAGGTAACGCGCAGACAGTGG 2040  
DB 279641 TAAAGAAATTAATGCGCTGATCAATTCGATATGAAACAACGAGGTAACGCGCAGACAGTGG 279582

QY 2041 GTGGCGGAAACCTTCAGGCGCGTAAGCAAGCTCTGGCGGATGTGATGACGAGCTCC 2100  
DB 279581 GTGGCGGAAACCTTCAGGCGCGTAAGCAAGCTCTGGCGGATGTGATGACGAGCTCC 279522  
QY 2101 GCCGACCTTCATGCTCGGAATTC 2123  
DB 279521 GCCGACCTTCATGCTCGGAATTC 279499  
  
RESULT 3  
AA67755  
ID AA67755 standard; DNA; 1575 BP.  
XX  
AC AA67755;  
XX  
DT 11-APR-2001 (first entry)  
XX  
DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:25.  
XX  
KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;  
KW membrane construction and membrane transport protein; petroleum spill;  
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;  
KW identification; microorganism; fine chemical production; transformation;  
KW genome mapping; genetic engineering; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN W020100805-A2.  
XX  
PD 04-JAN-2001.  
XX  
PE 23-JUN-2000; 2000MO-IB00926.  
XX  
PR 25-JUN-1999; 99US-0141031.  
PR 08-JUL-1999; 99DE-1031454.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031563.  
PR 09-JUL-1999; 99DE-1032122.  
PR 09-JUL-1999; 99DE-1032124.  
PR 09-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032180.  
PR 09-JUL-1999; 99DE-1032182.  
PR 09-JUL-1999; 99DE-1032182.  
PR 09-JUL-1999; 99DE-1032190.  
PR 09-JUL-1999; 99DE-1032191.  
PR 09-JUL-1999; 99DE-1032209.  
PR 09-JUL-1999; 99DE-1032212.  
PR 09-JUL-1999; 99DE-1032227.  
PR 09-JUL-1999; 99DE-1032228.  
PR 09-JUL-1999; 99DE-1032228.  
PR 09-JUL-1999; 99DE-1032229.  
PR 09-JUL-1999; 99DE-1032230.  
PR 14-JUL-1999; 99DE-1032927.  
PR 14-JUL-1999; 99DE-1033005.  
PR 14-JUL-1999; 99DE-1033006.  
PR 27-AUG-1999; 99DE-1040764.  
PR 27-AUG-1999; 99DE-1040765.  
PR 27-AUG-1999; 99DE-1040766.  
PR 27-AUG-1999; 99DE-1040830.  
PR 27-AUG-1999; 99DE-1040831.  
PR 27-AUG-1999; 99DE-1040832.  
PR 27-AUG-1999; 99DE-1040833.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.  
PR 31-AUG-1999; 99DE-1041395.  
PR 03-SEP-1999; 99DE-1042077.  
PR 03-SEP-1999; 99DE-1042078.  
PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042088.  
XX  
PA (BAD ) BASF AG.  
XX  
XX Pompeius M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;

DR MPI, 2001-071486/08.  
 XX P-PSDB; AAB76522.  
 PT Corynebacterium glutamicum nucleic acids encoding membrane construction  
 PT and membrane transport proteins or their portions, useful for typing or  
 PT identifying C. glutamicum or related bacteria, and as markers for  
 PT transformation -  
 XX  
 XX  
 PS Claim 3; Page 173-175; 1119pp; English.  
 CC AAF67743 to AAF68082 encode the Corynebacterium glutamicum membrane  
 CC construction and membrane transport (MCT) proteins given in AAB76510 to  
 CC AAB76847. The MCT nucleic acids and proteins are useful in the  
 CC identification of microorganisms which can be used to produce fine  
 CC chemicals, for modulating fine chemical production in C. glutamicum or  
 CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or  
 CC identification of C. glutamicum or related bacteria, as reference points  
 CC for mapping C. glutamicum genome, and as markers for transformation.  
 CC AAF68082 and AAF68082 represent sequencing primers which are used in an  
 CC example from the present invention.  
 XX  
 SO Sequence 1575 BP; 309 A; 434 C; 474 G; 358 T; 0 other;  
 Query Match 74.2%; Score 1575; DB 22; Length 1575;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1575; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 429 CTAATTTGGGCTTACATCTTCCGCTCTAATATGATGACAGACATTCGAATTAATTA 488  
 DB 1 CTAATTTGGGCTTACATCTTCCGCTCTAATATGATGACAGACATTCGAATTAATTA 60  
 QY 489 ACAAGGCCATTTTGGCCGCGGAGAGAGGTTTCCGACATATGATGAGGAGATGAACA 548  
 DB 61 ACAAGGCCATTTTGGCCGCGGAGAGAGGTTTCCGACATATGATGAGGAGATGAACA 120  
 QY 549 CACTTCAGCATGACGCTCATAGACTCGGTTTGGACCCCTGACAGCTTCATTTCTGGAA 608  
 DB 121 CACTTCAGCATGACGCTCATAGACTCGGTTTGGACCCCTGACAGCTTCATTTCTGGAA 180  
 QY 609 TGAATCTCCCAATATGACAACTCAATTAAGCTATGACAGACCTTGGAGCGGCTCG 668  
 DB 181 TGAATCTCCCAATATGACAACTCAATTAAGCTATGACAGACCTTGGAGCGGCTCG 240  
 QY 669 AAGCAAGGCCAATGCGATGATCGGTAATTAAGAGAGGACCGTGGAGGAGATTC 728  
 DB 241 AAGCAAGGCCAATGCGATGATCGGTAATTAAGAGAGGACCGTGGAGGAGATTC 300  
 QY 729 GGTAGCCGTTATTTTGTCCGATTTTCTTCTCGGCGGTTCTTTGGGACGCTCGCTC 788  
 DB 301 GGTAGCCGTTATTTTGTCCGATTTTCTTCTCGGCGGTTCTTTGGGACGCTCGCTC 360  
 QY 789 GGTGGCATCATGAAGGCGATTACCGCGCCACAGAGCTGAAACCTCCACTGCTGCTC 848  
 DB 361 GGTGGCATCATGAAGGCGATTACCGCGCCACAGAGCTGAAACCTCCACTGCTGCTC 420  
 QY 849 CCTCTCTCCGCTGGTGGCGCGCATGACAGGAACAATGAGCTTTTGTCAATGATGCTC 908  
 DB 421 CCTCTCTCCGCTGGTGGCGCGCATGACAGGAACAATGAGCTTTTGTCAATGATGCTC 480  
 QY 909 CATTAACCGCGCTGTGACAGCGTCAACCGGAGCGGATTTGCGCTTCTGATATTTGCG 968  
 DB 481 CATTAACCGCGCTGTGACAGCGTCAACCGGAGCGGATTTGCGCTTCTGATATTTGCG 540  
 QY 969 CATTCACAGATGAGTGGCGCGCATGAGCTCTGTGGGCTTCATTTGGGACATTCACCTTTGCG 1028  
 DB 541 CATTCACAGATGAGTGGCGCGCATGAGCTCTGTGGGCTTCATTTGGGACATTCACCTTTGCG 600  
 QY 1029 GGAACCCGCGCGCGAGATAGTTTCTGGGCTCTGCGGAGGAGTTAAACCATGGGCA 1088  
 DB 601 GGAACCCGCGCGCGAGATAGTTTCTGGGCTCTGCGGAGGAGTTAAACCATGGGCA 660  
 QY 1089 TGCCTTCCAGACGGTGTGACAGACGCGGAGATTTGTGAAACCTGTGTGATGATGCG 1148  
 DB 1148 TGCCTTCCAGACGGTGTGACAGACGCGGAGATTTGTGAAACCTGTGTGATGATGCG 720

DB 661 TGCCTTCCAGACGGTGTGACAGACGCGGAGATTTGTGAAACCTGTGTGATGATGCG 720  
 QY 1149 AATTGTGCGCACTTCCCAATTTGCGTGCAGCGGTGGCAAAAACCTTCAAGSTTATTCAGCC 1208  
 DB 721 AATTGTGCGCACTTCCCAATTTGCGTGCAGCGGTGGCAAAAACCTTCAAGSTTATTCAGCC 780  
 QY 1209 GGTAGAGGCAACGATGCTTTTCTTCCAACTCTGCGCGTGGCACTTCCGGTATGGA 1268  
 DB 781 GGTAGAGGCAACGATGCTTTTCTTCCAACTCTGCGCGTGGCACTTCCGGTATGGA 840  
 QY 1269 GCGATTTGCGCTTCTGTGACCCGAGAGGCTTGAATTCGGGAGATTTATGAAACGTT 1328  
 DB 841 GCGATTTGCGCTTCTGTGACCCGAGAGGCTTGAATTCGGGAGATTTATGAAACGTT 900  
 QY 1329 GGGGGCAACGCTGCTCAAGCTTTCTGTGCGGCTGCTGCGATTTGAGCCGCGTGTGCG 1388  
 DB 901 GGGGGCAACGCTGCTCAAGCTTTCTGTGCGGCTGCTGCGATTTGAGCCGCGTGTGCG 960  
 QY 1389 CGTTGCCGCGGCGCATTCGGGGGCGGCGCGTGTGCTGAATTTGGGACAGGATCGCGCTT 1448  
 DB 961 CGTTGCCGCGGCGCATTCGGGGGCGGCGCGTGTGCTGAATTTGGGACAGGATCGCGCTT 1020  
 QY 1449 CACGCTTGGGCGCGAGAGCTGCTTTTGGCGTGTGAGCAATTTGCTGCGCGGAGCT 1508  
 DB 1021 CACGCTTGGGCGCGAGAGCTGCTTTTGGCGTGTGAGCAATTTGCTGCGCGGAGCT 1080  
 QY 1509 AACTCTGCGATCTGTCTCATATGACACCTTCGCGCGCGCAATTTGTGCGAGCGGCTGA 1568  
 DB 1081 AACTCTGCGATCTGTCTCATATGACACCTTCGCGCGCGCAATTTGTGCGAGCGGCTGA 1140  
 QY 1569 GGAGCTGCGCATTCGCAACTGATTTGGCGGACCTGTGCAAGCTTATCGACGCTCCCT 1628  
 DB 1141 GGAGCTGCGCATTCGCAACTGATTTGGCGGACCTGTGCAAGCTTATCGACGCTCCCT 1200  
 QY 1629 CCCCAACCGTTTCTGATATTTGTCAGGCGGTTGGCGGTGGCGGCTGCGCATGCTCC 1688  
 DB 1201 CCCCAACCGTTTCTGATATTTGTCAGGCGGTTGGCGGTGGCGGCTGCGCATGCTCC 1260  
 QY 1689 GCGCATCTGCTCTACGCGGCGGCAAAACGCGTGTGCTGCGCATTTGCCACAGAGGCGC 1748  
 DB 1261 GCGCATCTGCTCTACGCGGCGGCAAAACGCGTGTGCTGCGCATTTGCCACAGAGGCGC 1320  
 QY 1749 CTGCGCATCTGCTTCCGCGGACCAACCAACGCGGCAAAATCATAGAGGCAAGAGGCT 1808  
 DB 1321 CTGCGCATCTGCTTCCGCGGACCAACCAACGCGGCAAAATCATAGAGGCAAGAGGCT 1380  
 QY 1809 GAGGCGGACGCACTTTTAAACCAAGGCTTATGACGCGGATGCTGCGGAAACCGAGCA 1868  
 DB 1381 GAGGCGGACGCACTTTTAAACCAAGGCTTATGACGCGGATGCTGCGGAAACCGAGCA 1440  
 QY 1869 CTTTGTGAAGAAATTTCTGCGCAATGACAGCGCTTCTCCGAATTTGATTAACATTC 1928  
 DB 1441 CTTTGTGAAGAAATTTCTGCGCAATGACAGCGCTTCTCCGAATTTGATTAACATTC 1500  
 QY 1929 GAGAGGCGGAGGAGCGGCACTGCTTCAACGATTTGAGGCTTTAGCGGAGTAAAGAA 1988  
 DB 1501 GAGAGGCGGAGGAGCGGCACTGCTTCAACGATTTGAGGCTTTAGCGGAGTAAAGAA 1560  
 QY 1989 ATTATGCGCTGATCA 2003  
 DB 1561 ATTATGCGCTGATCA 1575

RESULT 4  
 AAH65894  
 ID AAH65894 standard; DNA; 1473 BP.  
 XX  
 XX AAH65894;  
 XX AC  
 XX 26-SEP-2001 (first entry)  
 DT  
 XX  
 XX C glutamicum coding sequence fragment SEQ ID NO: 929.  
 DE  
 XX

KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KM organic acid synthesis; ds.  
XX Corynebacterium glutamicum.  
XX  
XX EPI108790-A2.  
XX  
XX 20-JUN-2001.  
XX  
XX 18-DEC-2000; 2000EP-0127688.  
XX  
XX 16-DEC-1999; 99JP-0377484.  
XX 07-APR-2000; 2000JP-0159162.  
XX 03-AUG-2000; 2000JP-0280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX WPI; 2001-376931/40.  
XX P-PSDB; AAG90675.  
XX  
XX Novel polynucleotides derived from Corynebacter bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT  
XX  
XX  
XX Claim 8; SEQ ID NO: 929; 246bp + Sequence listing; English.  
XX  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacter bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacter bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Corynebacter bacterium, and identifying a homolog of a gene derived  
CC from corynebacter bacterium. Corynebacter bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
XX  
SQ Sequence 1473 BP; 275 A; 414 C; 456 G; 328 T; 0 other;  
  
Query Match 69.4%; Score 1473; DB 22; Length 1473;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 508 GTGAGAGCGTTTCCGACATATGTTGGGCGATGGAACACATTGACGTTGACGTC 567  
DB 1 GTGAGAGCGTTTCCGACATATGTTGGGCGATGGAACACATTGACGTTGACGTC 60  
  
QY 568 ATGAGCTCGTTTGGACCTGACAGCTTCATTTCTTGAATGAACTCCCAATATGAC 627  
DB 61 ATGAGCTCGTTTGGACCTGACAGCTTCATTTCTTGAATGAACTCCCAATATGAC 120  
  
QY 628 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTGAGACCAAGGCCAATGCGAT 687  
DB 121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTGAGACCAAGGCCAATGCGAT 180  
  
QY 688 GAATCGGTAATTAATCGAGAGGACCGTGGAGGGCATTCGGTAAAGCCGTAATTTTGTCC 747  
DB 181 GAATCGGTAATTAATCGAGAGGACCGTGGAGGGCATTCGGTAAAGCCGTAATTTTGTCC 240  
  
QY 748 GATTTTCTCTCTCGGCGGTTCTTTGGGCAAGGCTCGGCTGCGATCATGATGAAGCG 807  
DB 241 GATTTTCTCTCTCGGCGGTTCTTTGGGCAAGGCTCGGCTGCGATCATGATGAAGCG 300  
  
QY 808 ATTCAACCGCGCACAGAGCTGAACTCCCACTGCTGCTCCCTGCTTCGGGTGTCG 867  
DB 301 ATTCAACCGCGCACAGAGCTGAACTCCCACTGCTGCTCCCTGCTTCGGGTGTCG 360

QY 868 CGCATGAGAGAGCAATTCAGCTTTTGTATGATGTTGTCATTAACCGCGCTGTGAC 927  
DB 361 CGCATGAGAGAGCAATTCAGCTTTTGTATGATGTTGTCATTAACCGCGCTGTGAC 420  
  
QY 928 CGTACCGGAGAGGCGCATTTGGCGTTCCGTTGATTTGGCAATCCAGATGGGTGGC 987  
DB 421 CGTACCGGAGAGGCGCATTTGGCGTTCCGTTGATTTGGCAATCCAGATGGGTGGC 480  
  
QY 988 GCCATGCGCTCGTGGGGTATCTGGGCATCTCACTTTTTCGGAACCCGCGCGCAGATA 1047  
DB 481 GCCATGCGCTCGTGGGGTATCTGGGCATCTCACTTTTTCGGAACCCGCGCGCAGATA 540  
  
QY 1048 GGTTCCTGAGTCTCTCGCGTGTGGAGTTAACCACTGGGCGATCGCTTCCAGACGTTGTG 1107  
DB 541 GGTTCCTGAGTCTCTCGCGTGTGGAGTTAACCACTGGGCGATCGCTTCCAGACGTTGTG 600  
  
QY 1108 CAGCAGCGGAGAGATTTGGTGAATACTGGTGTGATGATGAATTTGTGCGCACTCCAA 1167  
DB 601 CAGCAGCGGAGAGATTTGGTGAATACTGGTGTGATGATGAATTTGTGCGCACTCCAA 660  
  
QY 1168 TTGCGTCAAGCGGTGGCAAAACCTCAAGTTATTCAGCGGTGAGAGGCAACGATCGT 1227  
DB 661 TTGCGTCAAGCGGTGGCAAAACCTCAAGTTATTCAGCGGTGAGAGGCAACGATCGT 720  
  
QY 1228 TTTTTCCTCAACACTCTCGCGGTGGACATTCGCGGTGATGAGAGCGATGGCGCTTCTCGT 1287  
DB 721 TTTTTCCTCAACACTCTCGCGGTGGACATTCGCGGTGATGAGAGCGATGGCGCTTCTCGT 780  
  
QY 1288 GACCCGAGAGGCTGGATTCGGGAGATATGAAACGTTGGGGGCGACGTCGTCAG 1347  
DB 781 GACCCGAGAGGCTGGATTCGGGAGATATGAAACGTTGGGGGCGACGTCGTCAG 840  
  
QY 1348 CTTTCTGTGCGCGTGTGCGATTAAGCCGCGTGTGCGTTCCTGCGCGCATC 1407  
DB 841 CTTTCTGTGCGCGTGTGCGATTAAGCCGCGTGTGCGTTCCTGCGCGCATC 900  
  
QY 1408 GGGGGCGCGCGGTGGTGTGATTTGGGCAAGATTCGCGCTTCACGTTGGGCGCGAGAG 1467  
DB 901 GGGGGCGCGCGGTGGTGTGATTTGGGCAAGATTCGCGCTTCACGTTGGGCGCGAGAG 960  
  
QY 1468 CTGCGTTTTCGCGGTGTGATTTGCGTGGCGCGAGATTAACCTGCGCATCGTGTCC 1527  
DB 961 CTGCGTTTTCGCGGTGTGATTTGCGTGGCGCGAGATTAACCTGCGCATCGTGTCC 1020  
  
QY 1528 ATCATGACACTTCGCGCGCGAATTGTGCGAGCGGCTGAGAGGCTCGGCATCGCAAGC 1587  
DB 1021 ATCATGACACTTCGCGCGCGAATTGTGCGAGCGGCTGAGAGGCTCGGCATCGCAAGC 1080  
  
QY 1588 TCGATTGCGGCGACCTTGTCCAGAGCTTATCGAGCTCCCTCCCAACCGTTTGGTCAAT 1647  
DB 1081 TCGATTGCGGCGACCTTGTTCAGAGCTTATCGAGCTCCCTCCCAACCGTTTGGTCAAT 1140  
  
QY 1648 ATTGTGAGGGCGTTTGGCGGTGCGCGCTGCGCATGCTGCTGCTAGCGG 1707  
DB 1141 ATTGTGAGGGCGTTTGGCGGTGCGCGCTGCGCATGCTGCTGCTAGCGG 1200  
  
QY 1708 GCCGAAAACGCGTGTGCTGTCGATTCGACAGAGGCGCTCGCATCTTTCCGC 1767  
DB 1201 GCCGAAAACGCGTGTGCTGTCGATTCGACAGAGGCGCTCGCATCTTTCCGC 1260  
  
QY 1768 GACACCAACAGCGCGCGGAAATCATAGAGCGCAAGGCGTGAAGGCGACGCACTTTTA 1827  
DB 1261 GACACCAACAGCGCGCGGAAATCATAGAGCGCAAGGCGTGAAGGCGACGCACTTTTA 1320  
  
QY 1828 AGCCAAAGGCGTTATCGACGGGATCGTGCAGAAACGAGACACTTTTGAAGAAATTC 1887  
DB 1321 AGCCAAAGGCGTTATCGACGGGATCGTGCAGAAACGAGACACTTTTGAAGAAATTC 1380  
  
QY 1888 GGCACATACGACAGCGCTTCTCGAATTGATTAACATTCGAGAGGCGGAGCGCGAC 1947  
DB 1381 GGCACATACGACAGCGCTTCTCGAATTGATTAACATTCGAGAGGCGGAGCGCGAC 1440





```

Db      1261 GACACCAACGACGGCGGAAATCATPAGAGCAGAAAGGGTGCAGCGCAGCACTTTTA 1320
Qy      1828 AGCCAAAGGCTTATGACGGGATTCGTCCGCCAAACCGAGCACTTTGTTGAAGAAATTTCTC 1887
Db      1321 AGCCAAAGGCTTATGACGGGATTCGTCCGCCAAACCGAGCACTTTGTTGAAGAAATTTCTC 1380
Qy      1888 GGCACATCAGCAAGCCCTCTCCGAATTGATTAACAATCCGAGAGAGGCGGAGCGGAC 1947
Db      1381 GGCACATCAGCAAGCCCTCTCCGAATTGATTAACAATCCGAGAGAGGCGGAGCGGAC 1440
Qy      1948 AGTCGCTTACACGATTTGAGCGTTAGCGCAG 1980
Db      1441 AGTCGCTTACACGATTTGAGCGTTAGCGCAG 1473

```

RESULT 6  
AAH51981  
ID AAH51981 standard; DNA; 1488 BP.

```

AC      AAH51981;
XX      DT      04-SEP-2001 (first entry)
XX      DE      Mycobacterium tuberculosis potential drug target gene SEQ ID 35.
XX      KW      Drug target; growth; organism viability; characterisation; ds.
XX      OS      Mycobacterium tuberculosis.
XX      PN      WO200135317-A1.
XX      PD      17-MAY-2001.
XX      PF      13-NOV-2000; 2000MO-US31152.
XX      PR      12-NOV-1999; 99US-0165086.
XX      PR      12-NOV-1999; 99US-0165124.
XX      PR      01-FEB-2000; 2000US-0179531.
XX      PA      (BEC) UNIV CALIFORNIA.
XX      PI      Eisenberg D, Rotstein SH, Marcotte EM;
XX      DR      WPI: 2001-329193/34.
XX      DR      P-PSDB; AAG81130.

```

```

XX      PT      Identifying nucleotide or polypeptide sequence for use as drug target,
XX      PT      involves providing algorithm that analyzes a functional relationship
XX      PT      between nucleotide or polypeptide sequences, and comparing the
XX      PT      sequences -
XX      PS      Disclosure; Page 74; 207pp; English.
XX      CC      This invention relates to a method for identifying a nucleotide or
XX      CC      polypeptide sequence that may be a drug target, or essential for growth
XX      CC      or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
XX      CC      represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
XX      CC      tuberculosis proteins which are potential drug targets. The DNA and
XX      CC      protein sequences are used to illustrate the method of the invention. The
XX      CC      method involves providing an unknown nucleotide or polypeptide sequences,
XX      CC      and comparing it to a number of sequences along with at least one
XX      CC      algorithm capable of analysing a functional relationship between
XX      CC      nucleotide and polypeptide sequences. The method is useful for
XX      CC      characterising the function of nucleic acids and polypeptides that may be
XX      CC      useful as a target for a drug or essential for the growth or viability of
XX      CC      an organism.
XX      SQ      Sequence 1488 BP; 220 A; 509 C; 524 G; 235 T; 0 other;

```

```

Query Match      14.7%; Score 312.6; DB 22; Length 1488;
Best Local Similarity 54.0%; Pred. No. 1.1e-84;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

```

```

Qy      572 ACTGGTTTGGACCTGACAGCTTCAATTTCTTGAATGAACCTCCCAATATGACAAACC 631
Db      32 ACAGGCTGCTAGACCCGGGAAATCTTTGCTCAGCTGGATTAAGCAAGCCGCTGGCGTCCGG 91
Qy      632 TCAATCAAGGCTATGACAGAACCTTGAAGCGGGCTCGAAGCAAGGCCAAATGCGATGAAT 691
Db      92 TAGCCGACTCCTATGACCGGAGCTGGCCGCGCTCGGGCGGCGACACCGGCGGACGAAT 151
Qy      692 CCGTAAATTACTGAGAAAGGACACCTGAGAGGCAATCCGGTAAAGCCGTTATTTGTCGAT 751
Db      152 CGGTGACAGCCGGTGAAGAGACGCTAATTCGAGCGGCGGGTGGCCGTGTGCTGTAAGT 211
Qy      752 TTTCCTTCTCTGGAGGTTCTTTTGGGCAAGGTCGGTGGTGGTGGCATGATGAAGCGATTC 811
Db      212 TCGACTTCTGAGCGGCTTCGATTTGGGTGGCAGCGCCGACGATACCGCCGCGCTCG 271
Qy      812 ACCGCGCACAGACGTGAAACTCCCACTGCTGTCTCCCTGCTTCGGTGTGCGGCA 871
Db      272 AGCGGGGACCGCGGACCGGCTGCGCTCACTGGGCTCACCAAGCTCGGAGGCACTCCGCA 331
Qy      872 TGCAGAAAGACATGAGCTTTTGTCAATGATGATCATTAACCGCGGCTGTGAGGCTC 931
Db      332 TGCAGAAAGGACCGGTGCGCTTCTGCAAGATGTAAGATCGCTGCGGCATCCAGCTGC 391
Qy      932 ACCGCGAGCGCATTTGCGGCTTCTGCTGATTTTGGCAATCCACGATGAGGCGGCA 991
Db      392 ACAACGAGCGCGGCTGCTTCTTCTGCTGATTTTGGCAATCCACGATGAGGCGGCA 451
Qy      992 TGGCTGTGGGGTTCATCTGGGATCTCACTTTTGGGAAACCGGCGGCGAATAGATT 1051
Db      452 TGGCTGTGGGGTTCATCTGGGATCTCACTTTTGGGAAACCGGCGGCGAATAGATT 511
Qy      1052 TCTTGGGTCTCTGGGTGTGATGATTAACAAGGATGAGGCTTCCAGAGGTGTGAGC 1111
Db      512 TTTGGAAGCAACCGGCTTATGATGCTCTATGAGCAACCCCTTCCATCCGCGCTCAAA 571
Qy      1112 AGCGGAGAAATTTGTGTAACCTGTGATGATGATTTGTGTCGCACTCCCAATTGC 1171
Db      572 CCGCGAGAAATCTACGCGCGGATGAGATCATCAAGCGCGTCTGTGCACTGAGCGGCTAC 631
Qy      1172 GTGAGCGGTGCAAAAACCTCAAGGTTATTAGCCGGTATGAGGCAAGGATGTTT 1231
Db      632 GACCGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 691
Qy      1232 CTCGAACAATCTCTGGGTGTGCA---CTTCCGATGATGAGGCGCATTTGCCGTTCTGTG 1288
Db      692 CGCAGACGCGCGCGCGCGTACCCGATGTGCCACAGTGGGACTCGTGTGATGATGATGATG 751
Qy      1289 ACCGCGAGAGCTGGAATCGGGAGATTATGAAAGCTTGGGGGAGAGAGCTGCTCAAGC 1348
Db      752 GCGCGACCGCGCGCGCGCTCAGCGACTACTGCGACAGCGCGCGACCGAGCGGATGTT 811
Qy      1349 TTTCTGTGCGCGCTGTGCGGCAATGAGCCCGGCTGTGCGGCTTGCCTGTGCGCGCATTC 1408
Db      812 TGTGAGGAACCGATCAAGAGGAGAGCG---GGACCAAGCTGTGCGGCTGCGCGGCTTTC 868
Qy      1409 GGGCGCGCGCGGTGTGTGATTTGGGAG-----GATGCGCGCTTCAAGC 1453
Db      869 GCGCGCAACCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 928
Qy      1454 TTTGGCGCGAGAGCTGCGGTTTGGCGGTGTGCGGCTTGTGCGGCGCGGAGTAAAC 1513
Db      929 TGGGCGCGCGCTGCTTACGAGAACCGGACCGGAGTGGGCTTGGCGCGGAGCTGTGCC 988
Qy      1514 TGGCGATGCTGTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1573
Db      989 TGGCGCTGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1048
Qy      1574 TGGCGATGCGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1633
Db      1049 GCGGAGCTGGCGCGGACAGATGCGGATGCTGTGCGGAGCTGTGATGATGATGATGATGATG 1108

```

Qy 1634 CGTTTCGATTAATTGTCAGGCGGTGCGCGTGGCCGATCTGCCGCG 1693  
Db 1109 CGTGTGATCTGCTGGGCGAGGAGGCGCGCGCGCTGGGATGTTGCCCGG 1168  
Qy 1694 ATCTGTCTACGCGCGGAAACGCGTGTCTGTCGATTCGACGAGGCGCTCG 1753  
Db 1169 ACCGGGTGTGCGCCGATCCGCGGTGCGCGCGCTGCTGCTCCGAGGAGCGAGC 1228  
Qy 1754 CCATCTCTTCGCGGACCAACGCGCGGAAATCATAGAGGAGGAGGCGTGCAG 1813  
Db 1229 CGATGCTGTTCCGAGACACTGCTCATGCGCGGAACTGCTCCGCGGACATCCGT 1288  
Qy 1814 CGCAGCACTTTTAAAGCCAAAGGCTTATCGAGGATCGTCCGCAACCGACACTT 1873  
Db 1289 CGCGGACCTACTGAGTGGGGATTTGTGACACATCTGTGCGAGTACCCCGAGCGC 1348  
Qy 1874 TTGAAGA 1880  
Db 1349 CAGAGCA 1355

## RESULT 7

AA199683/C  
ID AA199683 standard; DNA; 4403765 BP.

AA199683;

15-JAN-2002 (first entry)

Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.

Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
variation; epidemiology; patient treatment; epidemic monitoring; ds.

Mycobacterium tuberculosis.

US6294328-B1.

25-SEP-2001.

24-JUN-1998; 98US-0103840.

24-JUN-1998; 98US-0103840.

(GENO-) INST GENOMIC RES.

Fleischmann RD, White OR, Fraser CM, Venter JC;

WPI; 2001-647261/74.

Evaluating strain variation of Mycobacterium tuberculosis, comprises  
determining the nucleotide sequence of the strain at positions in the  
genome corresponding to positions where M. tuberculosis strains CDC  
1551 and H37Rv differ

Claim 4; SEQ ID NO 2; 3bp + Sequence Listing; English.

The invention relates to evaluating strain variation within and between  
different populations of the tuberculosis bacterial pathogen.

Mycobacterium tuberculosis or related Mycobacterium by determining the  
sequence of the genome that correspond to positions that differ in the  
nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and  
H37Rv (AA199682). The method is useful for evaluating strain variation of  
M. tuberculosis and has valuable application in the fields of  
tuberculosis genetics, epidemiology, patient treatment and epidemic  
monitoring.

Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from USPTO  
at [seqdata.uspto.gov/sequence.html?DocID=6294328B1](http://seqdata.uspto.gov/sequence.html?DocID=6294328B1).

Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 14.7%; Score 312.6; DB 22; Length 4403765;  
Best Local Similarity 54.0%; Pred No. 1.4e-82;  
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

Qy 572 ACTGCTTTTGGACCTTACAGCTTCTTCTTGGAAATGAACCTCCAAATATGCAAC 631  
Db 1008148 ACGGCTGTAGACCGGGATCTTTTCTGACGTGGATAGGAGCGCTGGCGGTGCCG 1008089  
Qy 632 TCAATCAAGGCTATTCAGAGACCTTGGAGCGGCTTCGAAAGCAAGCCAAATCCATGAT 691  
Db 1008088 TAGCCGACTCTATGCGGGAGAGCTGCGCGCGCTCCGCGGCAACCGCGGAGCGAT 1008029  
Qy 692 CGTAAATTAAGTGAAGAGCAACCGTGAAGGCGATTCGGTATGCGGTATTTTGTCCGAT 751  
Db 1008028 CGGTCAACACCGGTGAGGAGCGCTATTCGAGCGCGGCGGTGCGGTGCGCTGTGAGT 1007969  
Qy 752 TTTCTTTCCTGCGCGGTCTTTTGGGACAGTGTGCGTCCGATCGCATGAAAGCGATTC 811  
Db 1007968 TCGACTTCTGCGCGGTCTGATTTGGGAGTGGAGCGCGCGGAGATCAACCGCGCGT 1007909  
Qy 812 ACCGCGCCACAGAGCTGAATCCCATCTGCTGCTCCCTGCTCCGCTGTCGCGCA 871  
Db 1007908 AGCGGCGCACCGCGAGCGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1007849  
Qy 872 TGCAGAGACATCGAGCTTTGTCATGATGATGTCATTAACCGCGGCTGTCAGCGCT 931  
Db 1007848 TGCAGAGAGCAACGCTGCTGCTTCTGAGATGATGATGATGATGATGATGATGATGAT 1007789  
Qy 932 ACCGCGAGCGCATTTGCGCTTCTGCTGATTTTGGCAATCCACAGATGGGTCGCGCA 991  
Db 1007788 ACAACAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1007729  
Qy 992 TGCGCTGTGGGCTTCAATCTGGCATCTCACTTTTGGGAAACCGCGCGCGCATAGATT 1051  
Db 1007728 TCGGCTGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1007669  
Qy 1052 TCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1111  
Db 1007668 TCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1007609  
Qy 1112 AGCGGAGATTTTGGTGAATACTGATGATGATGATGATGATGATGATGATGATGAT 1171  
Db 1007608 CGCGGAGATCTACGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1007549  
Qy 1172 GTGAGCGGTGCAAAAACCTCAAGGTTTCAAGCGGTGAGGAGCAACGATGCTTTT 1231  
Db 1007548 GACCGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1007489  
Qy 1232 CTCGAACAATCTGCGGTGCA---CTTCCGATGAGAGCGATTCGCGTCTGCTG 1288  
Db 1007488 CGCAGAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1007429  
Qy 1289 ACCCGAGAGCTGGAATCGGAGATTAATGAAAGTTGGGGGAGACGCTGCAAGC 1348  
Db 1007428 GCGCGAGACCGCGCGGCTGATGAGAGCTGCTGCAACAGCGCGCAACCGAGGTTGT 1007369  
Qy 1349 TTTCTGTGCGCTGCTGCGCATTTGAGCCGCTGCTGCGGTTGCTGCGCGCATG 1408  
Db 1007368 TGTGAGAACTGATCAAGAGGAGCG---GCAACAGCTGCTGCGCTGCGCTTGG 1007312  
Qy 1409 GGGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1453  
Db 1007311 GCGGCAACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1007252  
Qy 1454 TTGGCGCGCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1513  
Db 1007251 TCGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1007192  
Qy 1514 TGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1573  
Db 1007191 TGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1007132  
Qy 1574 TCGGCAATCGCAAGCTGATTTGCGCGCACTTTGTCAGGCTTATTCAGAGCTTCCCTCCCA 1633

Db 1007131 GCGGGCTGGCGGCGAGATTCGGCATTCCTGCGCGAGACTGTCTACGCTGGATACCCCGA 1007072  
Qy 1634 CCGTTTCGCTCATTTATTTGTGAAGGCGTTGGCGGTGGCGCGCTGCATTCCTGCGCGCG 1693  
Db 1007071 CCGTGTGCATCTCTGTGGGCGAGCGGCGAGCGCGCGCGCTGTGCGATGTTGCCCGCG 1007012  
Qy 1694 ATCTGCTAGCGCGGCGGAAACCGGTGCTGCTCCGCTGCATTCGACCAAGGCGCGCTCGG 1753  
Db 1007011 ACCGGTCTGCGCGCGCATTCACAGGCTGCGCGCGCTGCTGCTGCTGCGAGGAGCGAGCG 1006952  
Qy 1754 CCATCTCTTCGCGCGACCAACCAACCGCGGAAATCATAGAGCGAGCGCGTGCAGG 1813  
Db 1006951 CGATGCTGTTCCGAGACACTGCTCATTCGCGCGAACTGCTGCGCGCGCAAGGCGATCCGCT 1006892  
Qy 1814 CGCAGCGACTTTTAAGCCAAAGGCTTATTCAGCGGATGCTGCGCGAAACCGAGCATTTTG 1873  
Db 1006891 CGGCGGACTACTAGATTCGGGGAATTCGACACCATGCTGCGGAGTACCCCGACGCGCG 1006832  
Qy 1874 TTGAAGA 1880  
Db 1006831 CAGACGA 1006825

RESULT 8  
AI199682/c  
ID AI199682 standard; DNA; 4411529 BP.  
XX AI199682;  
AC AI199682;  
XX 15-JAN-2002 (first entry)  
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.  
XX  
XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
XX variation; epidemiology; patient treatment; epidemic monitoring; ds.  
XX Mycobacterium tuberculosis.  
XX  
XX US6294328-B1.  
XX  
XX 25-SEP-2001.  
XX  
XX 24-JUN-1998; 98US-0103840.  
XX  
XX 24-JUN-1998; 98US-0103840.  
XX  
XX (GENO-) INST GENOMIC RES.  
XX  
XX Fleischmann RD, White OR, Fraser CW, Venter JC;  
XX WPI; 2001-647261/74.  
XX  
XX  
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises  
XX determining the nucleotide sequence of the strain at positions in the  
XX genome corresponding to positions where M. tuberculosis strains CDC  
XX 1551 and H37Rv differ -  
XX  
XX  
XX Claim 3; SEQ ID NO 1; 3bp + Sequence Listing; English.  
XX  
XX The invention relates to evaluating strain variation within and between  
XX different populations of the tuberculosis bacterial pathogen,  
XX Mycobacterium tuberculosis or related Mycobacterium by determining the  
XX nucleotide sequence of the first strain at positions in the complete  
XX sequence of the genome that correspond to positions that differ in the  
XX nucleotide sequences of M. tuberculosis strains CDC 1551 (AI199683) and  
XX H37Rv (AI199682). The method is useful for evaluating strain variation of  
XX M. tuberculosis and has valuable application in the fields of  
XX tuberculosis genetics, epidemiology, patient treatment and epidemic  
XX monitoring.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from USPTO  
XX at seqdata.uspto.gov/sequence.html?docID=6294328B1.

XX  
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;  
Query Match 14.7%; Score 312.6; DB 22; Length 4411529;  
Best Local Similarity 54.0%; Pred. No. 1,4e-82;  
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;  
Qy 572 ACTGGTTTGGACCCCTGACAGCTTCACTTCTTGGAAATGAAATCCCAATATGACAAAC 631  
Db 1008147 ACCGGGTCTAGACCCGGGATCTTTGCTCAGCTGGATAGCAGCCGCTGCGCGCG 1008088  
Qy 632 TCAATCAAGCTATGACAGAGACCTTGGAGCGGGCTGGAAGCAAGCCAAATGCAATAT 691  
Db 1008087 TAGCCGACTCTATGCGCGGAGAGCTGCGCGCGCTGCGCGCGCGCGCGAGCAAT 1008028  
Qy 692 CGGTATTAATGAGAGAGCAACCGTGAAGGCAATTCGATAGCGCTTATTTTTCGATT 751  
Db 1008027 CGGTGACAGACCGGTGAGAGGAGCAGGTATTCGGGCGCGGGGTGCGCTGTGAGT 1007968  
Qy 752 TTTCTCTCTCGCGCGCTTCTTTGGGACCGGTGCGCTGCTGCGCTATCATGAAGCGATTC 811  
Db 1007967 TCGACTTCTGCGCGGCTTCGATTTGGGATGCGACCGCGCGCAAGATCACCGCGCTCG 1007908  
Qy 812 ACCGCGCAGACAGCTGAACCTCCCACTGCTGCTTCCCTGCTTCCGCTGCTGCGCGCA 871  
Db 1007907 AGCGGCGCACCGCGCGAGCGGCTGCGCTACTGACGCTACCAAGCTCGGAGGCAACCGCA 1007848  
Qy 872 TGCAGAAAGCAATCGAGCTTTTGTCAATGATGATTCATTAACCGCGCTGCGAGCTC 931  
Db 1007847 TGCAGAAAGGCAAGTGCCTTTCTGAGATGATGAAATCGCTGCGGCGCATCAGCTGC 1007788  
Qy 932 ACCGCGAGGCGCATTTGCGCTCTGCTGATTTTGGCAATCCACAGATGGTGGCGCCA 991  
Db 1007787 ACAACAGGCGCGCTGCGCTTACTGCTGATTTTGGCGCATTCGACCAAGGAGGTTT 1007728  
Qy 992 TGGCTTGTGGGTTCACTGCGCATCTCACTTTGCGGAACCGCGCGCGCATATAGTT 1051  
Db 1007727 TCGGTGCTGGGCTCGCTGGGCACTCAACCTGCGCAAGCGCGCGCTGATTCGCT 1007668  
Qy 1052 TCTGAGTCTCTGCGCTGCTGAGTAAACAATGGGATGCGCTTCCAGACGCTGCGAGC 1111  
Db 1007667 TTTGAGGACCAACCGGCTTATGATGCTTCTATGCGACCCCTTCCATCGCGGCTCAAA 1007608  
Qy 1112 AGCGGAGATTTTGTGAAACATGCTGATTTGATGAAATGCTGCGCACTCCAAATTCG 1171  
Db 1007607 CCGCGGAATCTACCGCGGCAATGATCAACGCGCGCTGTTGCACTGCAACCGGCTAC 1007548  
Qy 1172 GTGAGCGGTGGCAAAACCTCAAGTTATTCAGCGCGTGAAGGCAACGATCGTTT 1231  
Db 1007547 GACGATGCTGATTCGCTGCTTACGCTGCTCATTCAGCGCTCCGAAACGCTTCCGGC 1007488  
Qy 1232 CTCCAACATCTCTGGGCTGCA---CTTCCGATGATGAGGCGATTCGCGTTCTGCTG 1288  
Db 1007487 CGGAGACGCGCGCGCGCTGACCGATGTCGCCAGTGGGCTGCTGCTGCTGCGCGC 1007428  
Qy 1289 ACCGCGAGAGCGCTGGAATGCGGAGATTAATGAAACGTTGGGCGAGAGCTGTCAAGC 1348  
Db 1007427 GGCAGACGCGCGCGGCTGAGGAGCTACTGCAACGCGCGCACGAGCGGCTGTGT 1007368  
Qy 1349 TTTCTGCTGCGCTGCTGCGCATTTAGCGCGGCTGCGCGCTTGGCGCGCATCG 1408  
Db 1007367 TGTCAAGAAACGATCAAGGGAAGCG---GCAACACGCTGCTGCGCGCTGCGCGCTTGG 1007311  
Qy 1409 GGGGCGGCGCGGTGCTGATTTGGGCGA-----GATCGCGCTTCAACCGC 1453  
Db 1007310 GCGGCGCAACCGAGGCTCTCGGCGAGAAAGGCGAGTGAAGCGCGGCGGGAAGCATG 1007251  
Qy 1454 TTGGGCGGAGAGCGCTGCTTGGCGGCTGCGGCAATTCGCTGCGGCGCGGAGCTAAAC 1513  
Db 1007250 TCGGCGCGCGCTGCTTACGGAAGCGCGAGCGGAGTGGGCTGCGCGCGAGCTGTGCC 1007191  
Qy 1514 TGGCGATGCTGCTCATCATGACACTCGCGCGCGCAATTTGCGAGCGGCTGAGAGAGC 1573







PI Fontana MR, Pizza M, Masignani V, Monaci E;  
 XX WPI; 2003-058415/05.  
 DR P-PSDB; ABP78030.  
 XX  
 PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a  
 PT medicament for treating or preventing *N. gonorrhoeae* infection -  
 XX  
 PS Disclosure; Page 377; 815pp; English.  
 XX  
 CC The present invention relates to proteins from *Neisseria gonorrhoeae*.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records AB237706-AB242016 represent nucleic acid  
 CC molecules of the invention.  
 XX  
 SQ Sequence 876 BP; 183 A; 216 C; 272 G; 205 T; 0 other;

Query Match 5.8%; Score 122.2; DB 25; Length 876;  
 Best Local Similarity 51.6%; Pred. No. 1.9e-26;  
 Matches 280; Conservative 0; Mismatches 263; Indels 0; Gaps 0;

QY 628 AACCTCAATCAAGGCTATGACAGACCTTGAGCGGCGTCAAGAGGCGCAATGCGAT 687  
 DB 274 AAAGACAGAAAATATTCGACCGCTTGAATGCGGACGACGAGCTGACCGGAAAGAT 333  
 QY 688 GAATCGTAATTAATCTGAGAGAGGACCCGTGAGGGGCAATCCGTAATTTTGTCC 747  
 DB 334 GACGCGCTGTGTGATGAAGGAGATGATGAACGGCTCCCGTGTGTCGCGGTTT 393  
 QY 748 GATTTTCTTCTCTGCGGCTTCTTTGGACAGGTCGCTCGGTGCGATCATGAAGGCG 807  
 DB 394 GAATTCCTCTTATGCGCGGCTTCGATGGGTGCTTGGGCGACGTTTCTTACAGGGT 453  
 QY 808 ATTCACCGCGCACAGAGCTGAATCCCACTGCTGTCTCCCTGCTTCCGGTGTGGC 867  
 DB 454 GTCCCGCTGCGGCTTCCGATTAATGCTGCTTGTCTGTGTGGCGGCTTCCGGCGCGC 513  
 QY 868 CGCATGACAGAAAGACATCGAGCTTTTGTCTATGATGTGTGTCATACCGCGCTGTGCA 927  
 DB 514 CGTATGACAGAGGGTGTAACTGCTGATGACAGATGAGAAACAGTGGCGGCTGAT 573  
 QY 928 CGTACCGCGAGCGCCATTTGCGCTTCTGCTGTATTTGGCGAATCCACGATGGGTGC 987  
 DB 574 TTGCTGACGAGAAAACGCTGCGCTTATATCGGTGTGACCGATCCGATATGGGCGGC 633  
 QY 988 GCATAGCGCTGCGGGTTCATCTGGGCACTCATCTTTTGGGGAACCGCGCGGAGATA 1047  
 DB 634 GTATGTGCAAGCTTTCGATTTTAAAGCGATGTGCTGCGAAGCGAGCGCTGATC 693  
 QY 1048 GGTTCCTGCGGTCTCGCGGTGTGAGTTAACCACTGGGAGTCGCTTCAAGCGGTGTG 1107  
 DB 694 GGTTCCTGCGGTCTCGCGGTGTGAGTTAACCACTGGGAGAACCGCTGCGGAAGGCTTC 753  
 QY 1108 CAGCAGCGCGGAGATTTGTGAAAACTGTGTGATTTGATGATTTGTGCGCACTCCAA 1167  
 DB 754 CAACGCGCGGAGTTTCTCTGAAAAAAGCGGATGACCAAGATTTGACCGCGCGCAT 813  
 QY 1168 TTG 1170  
 DB 814 ATG 816

RESULT 13  
 AAS53191  
 ID AAS53191 standard; DNA; 867 BP.  
 AC AAS53191;  
 XX  
 DT 13-FEB-2002 (first entry)

XX  
 DE Enterococcus faecalis DNA for cellular proliferation protein #619.  
 XX  
 KW Antisense; ds; prokaryotic cellular proliferation gene;  
 KW antibiotic; antibacterial; drug design.  
 XX  
 OS Enterococcus faecalis.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207127P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GT;  
 PI Yamamoto RT, Xu HH;  
 DR WPI; 2001-611495/70.  
 DR P-PSDB; AAU35332.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Claim 27; Seq ID No 6828; 51pp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence encodes an essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

SQ Sequence 867 BP; 266 A; 145 C; 209 G; 247 T; 0 other;

Query Match 5.7%; Score 121; DB 23; Length 867;  
 Best Local Similarity 53.0%; Pred. No. 4.3e-26;  
 Matches 259; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 685 GATGATTCGTAATTAATCTGAGAGGACCGCTGAGGGCAATCCGGTATTTTGG 744  
 DB 328 GATGATTCGTAATTAATCTGAGAGGAGCAATTAAGAGCAAGGTGTGCAATTTGATC 387  
 QY 745 TCCGATTTTCTCTCTCGGCGGTTCTTTGGGACAGGTCGCGTGGCGCATGATGAG 804  
 DB 388 ATGACGCTAATTTATCATGAGGAGATGAGAAAGATTTGTGTGAAAAATACACGC 447  
 QY 805 CGGATTCACCGCGCACAGAGCTGAATCCCACTGCTGTCTCCCTGCTTCGGGTGT 864  
 DB 448 TTGTTTGAAGCGGCGAGAGAAAGCAATTTACAGTATGATTTTCACTGCATCTGTGCT 507

QY 865 GCGGCGATGCGAAGCAATCGAGCTTTTGTATGATGATGTCCTCAATACCGGCGCTGTG 924  
DB 508 GCCCGTATGCAAGAAAGAAATTTTTCATTGATGCAAAATGCGAAAATTTTCGGCGCTTTG 567  
QY 925 GAGCGTCACCGCGAGGCGCATTTGCGCTTCCTGCTGATTTGGCGAATCCGCGATGGGT 984  
DB 568 CAACGCGATTAACAAGAGGCTTGCTGTATCTTACGCTATGACTGATCCAAAGCACTGGC 627  
QY 985 GGGCGCATGCGCTCGTGGGCTTATCTGGGCACTCACTTTGGGAGACCGCGCGCAG 1044  
DB 628 GGGTTACCGCAAGTTTGGATGATGATGCGATATATTTTGGCAGAGCTCAGAGTTTA 687  
QY 1045 ATAGTTTCTCTGGGCTCTGCGGCTGATGATTAACCACTGGGAGTGGCTTCCAGAGCT 1104  
DB 688 ATCGTTTGTCTGCGCGCGCTGTAATGTAACAACCAATTCGTCAAGAGTTGCCAGATGAT 747  
QY 1105 GTGCGACAGCGGAGAAATTTGTGTAACCTGTGTGATGATGAAATTTGTGCGCATCTC 1164  
DB 748 TTTCAAAAAGCGCAGTTTCTTTAGAACATGTTTGTGATCAATGTGTGCGAAGAAAT 807  
QY 1165 CAATTGCGT 1173  
DB 808 CTTTTCGCT 816

## RESULT 14

AA13395

ID AA13395 standard; DNA; 6021 BP.

XX AA13395;

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:458.

KM Enterococcus faecalis; contig; detection; Enterococcal infection;  
vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN W09850555-A2.

XX 12-NOV-1998.

PF 04-MAY-1998; 98WO-US08985.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

XX 16-MAY-1997; 97US-0046655.

(HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

XX WPI; 1999-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
- used to develop products for the detection of Enterococcus and for  
PT use in vaccines for prevention or attenuation of Enterococcus  
infection.

PS Claim 1; Page 1691-1694; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it  
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
CC AA13395 to AA13399 represent these nucleotide sequences which are  
CC primary nucleotide sequences, also known as contigs. The computer-based  
CC system can identify fragments of the Enterococcus faecalis genome with  
CC commercial importance. The products can be used to detect the presence  
CC of Enterococcus faecalis in samples. They can also be used for  
CC diagnosing Enterococcal infection in an animal and monitoring  
CC progression of disease, and for identifying agents which can be used to  
CC modulate the growth or pathogenicity of Enterococcus faecalis, orCC another related organism, in vivo or in vitro. In particular the  
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
CC can be used in vaccines to prevent or attenuate an Enterococcal  
CC infection.

SQ Sequence 6021 BP; 1920 A; 1007 C; 1415 G; 1673 T; 6 other;

Query Match 5.7%; Score 121; DB 20; Length 6021;  
Best Local Similarity 53.0%; Pred. No. 1,4e-25;  
Matches 259; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 685 GATGAATCGGTAATTTACTGGAAGAAAGCAACCGTAGGCGCAATTCGGTAGCCGTTATTTTG 744  
DB 5321 GATGAACGCTCTTAACGGAGAAAGCAACATTTGAAGACAAACGCTGTGCAATTTGAATC 5380  
QY 745 TCCGATTTTCTCTCTCGGCGGTTCTTTGGGACCGTCCGCTGGTCCGCAATCAAGAG 804  
DB 5381 ATGAGCGCTAATTTTATCATGGGCACTATGGAACGATTTGGTGAATAAAATCAGACGC 5440  
QY 805 GCGATTACCGCGCAAGAGCTGAAATCCACGCTGCTCTCCCTGCTTCCGCTTCCGCTGGT 864  
DB 5441 TTGTTGAGCGGCGCAGAAAGCAATTTACAGTATGATTTTCACTGATCTGCTGTGT 5500  
QY 865 GCGGCGATGCGAAGAAAGCAATTCAGCTTTTGTATGATGATGTCATTAACCGGCTGTG 924  
DB 5501 GCCGTAATGCAAGAAAGAAATTTTTCATTGATGCAAAATGCGAAATTTTCGGCGCTTTG 5560  
QY 925 CAGCGTCACCGCGAGGCGCATTTGCGCTTCTGCTGATTTTGGCGCAATCCAGATGGGT 984  
DB 5561 CAACGCGATTAACAAGAGGCTTGTCTGTATCTTACGATGATGATCAACAGCTGGC 5620  
QY 985 GGGCGCATGCGCTCGTGGGCTTATCTGGGCACTTCATCTTTTGGGAAACCGCGCGCAG 1044  
DB 5621 GGTGTACCGCAAGTTTGTGATGATGCGATATATTTTGGCAGAGCTCAGAGTTTA 5680  
QY 1045 ATAGTTTCTCTGGGCTCTGCGGCTGATGATTAACCACTGGGAGTCCGCTTCCAGAGCT 1104  
DB 5681 ATCGTTTGTCTGCGCGCGCTGTAATGTAACAACCAATTCGTCAAGAGTTGCCAGATGAT 5740  
QY 1105 GTGCGACAGCGGAGAAATTTGTGTAACCTGTGTGATGATGAAATTTGTGCGCATCTC 1164  
DB 5741 TTTCAAAAAGCGCAGTTTCTTTAGAACATGTTTGTGATCAATGTGTGCGAAGAAAT 5800  
QY 1165 CAATTGCGT 1173  
DB 5801 CTTTTCGCT 5809

## RESULT 15

ABS99190

ID ABS99190 standard; DNA; 6021 BP.

XX ABS99190;

DT 18-DEC-2002 (first entry)

DE Enterococcus faecalis contig sequence #458.

KM Computer readable medium; Enterococcus faecalis; microbe; growth;  
KM pathogenicity; vaccine; resistance; Enterococcal infection; commercial;  
KM therapeutic; industrial; fermenting; sugar source; metabolic; vaccine;  
KM biotech technology; antibacterial; modulator of nucleic acid expression;  
contig; ds.

OS Enterococcus faecalis.

XX US2002120116-A1.

XX 29-AUG-2002.

PF 04-MAY-1998; 98US-0070927.

XX 04-MAY-1998; 98US-0070927.

XX (KUNS/) KUNSCH C A.  
PA (DILL/) DILLON P J.  
PA (BARA/) BARASH S.  
XX

PI Kunsch CA, Dillon PJ, Barash S;  
XX

DR WPI, 2002-750065/81.  
XX

XX Computer readable medium having recorded on it a Enterococcus faecalis  
PT nucleotide sequence useful for detecting diseases related to  
PT Enterococcus infections in animals  
XX

PS Claim 1, Page -; 119pp; English.  
XX

CC The present invention relates to a new computer readable medium with an  
CC Enterococcus faecalis nucleotide sequence. The invention is useful to  
CC diagnose the presence of E. faecalis in a sample or determining the  
CC presence of a specific microbe in a sample. The invention is also useful  
CC for modulating the growth or pathogenicity of E. faecalis, in a vaccine  
CC to confer resistance to Enterococcal infection, for commercial  
CC therapeutic and industrial purposes, and for fermenting a particular  
CC sugar source or to produce a particular metabolite. The invention is  
CC useful for detecting diseases related to Enterococcus infections in  
CC animals, and for detecting E. faecalis using biochip technology. The  
CC present nucleic acid sequence represents an Enterococcus faecalis conj  
CC DNA sequence of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification but was obtained in electronic format directly  
CC from USPTO at <http://seqdata.uspto.gov>.  
XX

Sequence 6021 BP; 1920 A; 1007 C; 1415 G; 1673 T; 6 other;

Query Match 5.7%; Score 121; DB 24; Length 6021;  
Best Local Similarity 53.0%; Pred. No. 1.4e-25;  
Matches 259; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 685 GATGATGCGTAACTACTGAGAGAGACCGTGGAGGCGCATCCGGTAGCCGTTATTTTG 744  
DB 5321 GATGAAAGCTGTCTTAAAGGAGAGACAAATTTGAAGCAAGCTGTGCAATTGGAATC 5380  
QY 745 TCCGATTTTCTCTCTCGCGGTTCTTTGGACAGTCCGCTCGTGGCATATGAAG 804  
DB 5381 ATGACGCTAATTATATATGAGAGATGGAACGATTTGTGTGAATAAATCACAGC 5440  
QY 805 GCGATTACCGCGCCACAGAGCTGAACCTCCACTGCTGCTCCCTGCTTCCGCTGT 864  
DB 5441 TTGTTGAGCCGGCCAGAAAAGCATTTACAGTAGTGATTTTCACTGCATCTGTGT 5500  
QY 865 GCGGCATGCGAGAGACATGAGCTTTTGCATGATGAGTGCATTAACCGCGGCTG 924  
DB 5501 GCCCGTATGAAAGAGAAATTTTTCATTTGATGCAAAATGCGAAATTTGGCGCTTTG 5560  
QY 925 CAGCGTCAACCGGAGCGCATTTGCGGTTCTCTGTTATTTGCGCAATCCACAGATGGT 984  
DB 5561 CAACGCAATACAAAGCAGGCTTCTGCTATCTTAAGGTAATTCATGATCCAAAGACTGGC 5620  
QY 985 GCGCCATGCGCTCTGTGGGTTCACTGGGCACTCACTTTTGGGAAACCGCGCGCAG 1044  
DB 5621 GGTGTTACCGCAAGTTTTCGATGATGCGATATTTTGGCAGAGCCTCAGAGTTTA 5680  
QY 1045 ATAGGTTCTCTGCGCTCTCGGCTGATGAGTAAACCATGCGGATCGCTTCCAGACGT 1104  
DB 5681 ATCGGTTTCTGCGCGCGCTGTAAATGAAACAAAGATTCGTAAGAGTTGCCAGATGAT 5740  
QY 1105 GTGACAGAGCGAGAAATTTGTGAAAACGTGTTGATGATGATTTGTGCGCACATC 1164  
DB 5741 TTTCAAAAGCGCGAATTTCTTTAGAACATGTTTTTTGATGATCAGATTTGCCAAGAAT 5800  
QY 1165 CAATTGCGT 1173  
DB 5801 CTTTTCGT 5809

Search completed: November 12, 2003, 20:25:38  
Job time : 625.017 secs

**THIS PAGE BLANK (USPTO)**



9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjoftus@ictr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seq primer: M13 For

# FEATURES

source Location/Qualifiers

1..488  
/organism="Anopheles gambiae"  
/mol\_type="genomic DNA"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-133C23"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 137 a 86 c 109 g 156 t

## ORIGIN

Query Match 4.1%; Score 88; DB 28; Length 488;  
Best Local Similarity 49.0%; Pred. No. 3.4e-14;  
Matches 235; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

639 AGGCTATGAGACCTTGGAGCGGCTCGAGACGCAATCGATGATCGGTAT 698  
1 AGCTTATCTGACCCATTAAAGAGTAAAGGCAAGCTAAATTAACGATTCCTGAC 60  
639 TAATGAGAGGACACCGTGAAGGCAATCCGCTAGCCGTTATTTTGTCCGATTTTCTT 758  
61 TAATGCTGTGTAAGTAAAGCAAGCACTGGAATCTGTTGTTATGATGATGATTTCTTTT 120  
759 CCTGCGCGGTTCTTTGGGACGCGTCCGCTGCGTGCATGAAAGGCAATTCACCGGC 818  
121 TATCGAGGCTCATTAATGCTCTGTATAGGAGAAAAAATCATCCGCTATGCACTATGC 180  
819 CACAGAGCTGAAGTCCCACTGCTGCTCCCTGCTCCGCTGCGGTCGCGATCAGGA 878  
181 AATGAGCATTAATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 240  
879 AGACAATCGAGCTTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 938  
241 AGCAGCTTATTCATTAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 300  
939 GGGCGATTTCCGCTTCTGCTGATTTTGGGCAATCCCAAGATGGTGGCGCATGCGCTC 998  
301 CAAGGCTTACCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 360  
999 GTGGGATTCATGCGGATCTCACTTTTGGGACCCGCGCGCGATGATGATGATGATGATG 1058  
361 TTTTGGATGATCTGCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 420  
1059 TCCTGCGGCTGATGATTAACCACTGGGATGCGCTTCAGACGATGATGATGATGATGAT 1118  
421 TCCTGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 480

RESULT 2  
BU654446 707 bp mRNA linear EST 30-SEP-2002  
LOCUS 112131P06.v1 C. reinhardtii CC-1690 (mt-), CC-1691 (mt-), Gamete  
DEFINITION (normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA  
ACCESSION BU654446  
VERSION BU654446.1 GI:23366627  
KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadaceae; Chlamydomonas.

## REFERENCE

1 (bases 1 to 707)  
Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J.P., Shrago, J., Sillflow, C. and Stern, D.  
Analyses of the Chlamydomonas reinhardtii Genome: A Model, Cellular System for Analyzing Gene Function and Regulation in  
Vascular Plants. Project: 1112

JOURNAL Unpublished  
COMMENT Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.

## FEATURES

source Location/Qualifiers

1..707  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="21gr (CC-1690 wild type mt+) & 6145c (CC-1691 wild type mt-)"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690 (mt+), CC-1691 (mt-), Gamete (normalized), Lambda Zap II"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; Gamete library was constructed by John Zhao, Jeffrey McDermott, William J. Sneli and Huu Davies. Strain 21gr cells (CC-1690, mating type plus) and strain 6145c cells (CC-1691, mating type minus) that had been growing on a light-dark cycle (13:11 L/D) in R-medium (Seger and Granick) were separately transferred into nitrogen-free medium at 8 hours into the light period. PolyA mRNA was purified from each sample every 2 hours for the next 18 hours. The mRNA was pooled and used for cDNA synthesis. The cDNA was directionally cloned into Lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the Lambda Zap clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

BASE COUNT 127 a 241 c 227 g 112 t

## ORIGIN

Query Match 4.1%; Score 86.2; DB 13; Length 707;  
Best Local Similarity 50.6%; Pred. No. 1.4e-13;  
Matches 208; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

1449 CACGCTGGGCGGAGAGCTGCTTTGGCGCTGCGATTTCCGTGCGCGGAGCT 1508  
286 CATGCCCGGCGCAAGCGTACCGCAAGGCGCTGCGTTATGCGCCAGCGCAAGTT 345  
1509 AAACCTGCGATCGTGCATCATGACACCTCCGCGCGCAATTTGCGAGGCGGTGA 1568  
346 TGTCTGCGCATCATCTTCTGAGACAGCCCGAGGCTATGCGCGCAAGACCGGGA 405  
1569 GAGCTGGCATTCGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1628  
406 GAGCTGGCGGAGGCGGAGGCGATTCGCGTGAACCTCGGAGATGTTCCGCGTGT 465  
1629 CCCACCGTTTGGTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 1688  
466 GCCCATATCTCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 525  
1689 GCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1748  
526 GCGCAACGCACTGATCATGAGAGCGGTTTACTAGTGCGCTCGCCGAGGCTG 585  
1749 CTGCGCATCTCTTTCGCGGACCAACCAAGCCCGGCAATCATGAGGCAAGGCGT 1808



Db 586 CGCGCCATCTGTGAGAGCCGCTCTGCCCGGCGAGCCACTGAGCCCTGGCAT 645  
 Oy 1809 GCAGGCGCAGCACTTTAAGCCAGGCTTATCGAGGATGTCGCCGA 1859  
 Db 646 CACCTGCGCGAGCTGTGAAGTTCCGCTCATGACCAATGTCGCCGA 696

## RESULT 3

AM584783  
 LOCUS

DEFINITION 604 bp mRNA linear EST 07-SEP-2000

AM584783  
 library cDNA clone MHAM-7110, mRNA sequence.

ACCESSION AM584783  
 VERSION AM584783.1 GI:7261837

KEYWORDS Medicago truncatula/Gloms versifforme mixed EST library

SOURCE Medicago truncatula/Gloms versifforme mixed EST library

ORGANISM Medicago truncatula/Gloms versifforme mixed EST library

REFERENCE 1 (bases 1 to 604)  
 Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and Fraser, C.M.

AUTHORS Harrison, M.J., Liu, J., Peng, H., Gonzales, M., Ellis, L., Town, C.D.,  
 ESTs from roots of Medicago truncatula after colonization with

TITLE Gloms versifforme

JOURNAL Unpublished

COMMENT Contact: Harrison M.J.  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73401  
 Tel: 580-223-5810  
 Fax: 580-221-7380

Email: mjharrison@noble.org  
 Other name: MHAM-7b-805; Date: 3/14/00; Updated to the Database of

Expressed Sequence Tags (dbEST) on 04/27/00; More information is

available at 'http://chrysle.cam.ac.uk/medicago'.

Seq primer: 73.

FEATURES

source Location/Qualifiers

1..604  
 /organism="Medicago truncatula/Gloms versifforme mixed EST

library"

/mol\_type="mRNA"

/cultivar="Medicago truncatula genotype A17"

/db\_xref="taxon:119092"

/clone="MHAM-7110"

/tissue\_type="roots colonized with Gloms versifforme"

/dev\_stage="Roots harvested at 10, 17, 22, 31 and 38 days

post-inoculation with Gloms versifforme. The library was

made from a mixture of RNA from each of these stages."

/lab\_host="E. coli strain XL0LR"

/clone\_lib="MHAM"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI; cDNA was prepared from polyA+ enriched RNA from

roots harvested at 10, 17, 22, 31 and 38 days

post-inoculation with Gloms versifforme. The cDNA was

directionally ligated into the UniZap XR vector from

stratagene and packaged using Gigapack III Gold packaging

extracts. Plasmids containing cDNA inserts were excised

from the recombinant lambda-Zap phage using Ex-assist

helper phage and propagated in XL0LR cells."

BASE COUNT 179 a 95 c 136 g 194 t

ORIGIN

Query Match 4.0%; Score 84.2; DB 9; Length 604;

Best Local Similarity 48.8%; Pred. No. 56-13; Indels 3; Gaps 1;

Matches 258; Conservative 0; Mismatches 268;

634 AATCAAGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGCCAAATGCGATGATCG 693

13 AATGATGATTTACAAATCGCTTATCTTATCAAGACAGACCGATTAATGATCG 72

634 GTAATTACTGAGAGGACCGCTGAGGCGATTCGGTAGCCGTTATTTTGTCCGATTT 753

73 GTTCAACAGGACAGCTCAAGTAATGATTTCTGTAGCAATAGTATATGATTTT 132

Oy 754 TCCCTTCCTGCGCGCTTTCTTGGGACAGGTCGCGTGGTCCGATCATGAGGCGATTCC 813  
 Db 133 GAGTTATGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 192  
 Oy 814 CGCGCCACAGAGCTGAACTCCCACTGCTGCTCCCTGCTTCCGCTGCTGCTGCTGCT 873  
 Db 193 TATGCTACCAATCAACGTTTACCTCTTATTAATGATGATGATGATGATGATGATGAT 252

Oy 874 CAGAGAGCAATCGAGCTTTTGTCTATGATGATGATGATGATGATGATGATGATGATGAT 933  
 Db 253 CAAAGAGAGAGAGTTGAGCTTAATGCAAAATGCTAAATTTTCTTTTATATATATAT 312

Oy 934 CGCGAGCGCATTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 990  
 Db 313 CAAATCAATCAAAATGATTTTCTATGATGATGATGATGATGATGATGATGATGATGAT 372

Oy 991 ATGCGCTGCGGCTTCAATCGGAGCTCACTTTTGGGAGACCCGCGCGAGATAGT 1050  
 Db 373 ACAGCTAGTTTGGAGATGTTGGGGATATCAATTTTCCGGAACCGAGTCTTACATAGCA 432

Oy 1051 TTCTCGGCTCTCGCGTGTGAGATTACCACTGCGCATGCGCTTCCAGACGCTGCA 1110  
 Db 433 TTTCGAGGTAAAGATTAATGAGAAAGTTGAAGATTGAAGTCCGAAAGTATACAA 492

Oy 1111 CAGCGGAGAAATTTGGTAAAGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 1159  
 Db 493 TCGGCTGAATTTTATTTGAAAGAGGTGCTATTTGATGATGATGATGATGATGATGAT 541

RESULT 4  
 B1311746  
 LOCUS

DEFINITION EST5313496 GUSD Medicago truncatula cDNA clone pGSD15E20 5' end,  
 mRNA sequence.

ACCESSION B1311746  
 VERSION B1311746.1 GI:14986073

KEYWORDS EST.

ORGANISM Medicago truncatula (barrel medic)

REFERENCE 1 (bases 1 to 773)  
 Grusak, M.A., Samac, D.A., Town, C.D., Van Aken, S., Utterback, T., Cho

AUTHORS J., and Fraser, C.M.

TITLE ESTs from developing reproductive tissues of Medicago truncatula

JOURNAL Unpublished

COMMENT Contact: Michael A. Grusak  
 USDA/ARS Children's Nutrition Research Center  
 Baylor College of Medicine  
 1100 Bates Street, Houston, TX 77030-2600, USA  
 Tel: 713-798-7044  
 Fax: 713-798-7078  
 Email: mgrusak@bcm.tmc.edu

B398601e

TIGR sequence name: MTPA234TK

More information is available at: www.medicago.org

Seq primer: SKmod (CTA GAA CTA GTG GAT CC).

FEATURES

source Location/Qualifiers

1..773  
 /organism="Medicago truncatula"

/mol\_type="mRNA"

/cultivar="A17"

/db\_xref="taxon:3880"

/clone="pGSD15E20"

/tissue\_type="immature seeds"

/dev\_stage="Immature seeds, 11 to 19 days after

pollination"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI; Immature seeds, collected from pods ranging in age

from 11 to 19 days after pollination, were harvested from

greenhouse-grown plants. Seeds were removed and separated from pod walls and were immediately frozen in liquid nitrogen. Seeds throughout the age range were pooled for mRNA extraction. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the UniZap XR vector from Stratagene and packaged using GigaPack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-assist helper phage and propagated in XLOLR cells.

BASE COUNT 261 a 112 c 154 g 246 t  
ORIGIN

Query Match 3.9%; Score 82.6; DB 12; Length 773;  
Best Local Similarity 48.6%; Pred. No. 1.7e-12;  
Matches 257; Conservative 0; Mismatches 269; Indels 3; Gaps 1;

Qy 634 AATCAAGGCTATGACAGACCTTGGAGCGGCTCGAACCAAGGCCAATGGATGATCG 693  
Db 33 AATGATGATATCAAAATCGCTTATCTTATCAAGACAGAACCGATTACTGATCG 92  
Qy 694 GTAATTATGAGAGAGCAGCGTGGAGGCAATCCGTTAGCCGTTATTTTGTCCGATTT 753  
Db 93 GTTCAACAGGACAGGTCAGTAATGATGATCTCTGAGCAATGATGATATGATTT 152  
Qy 754 TCCCTTCTCGGCGCTTCTTGGGCAAGGTCGCGTGGGCGATGATGAGCGGATTCAC 813  
Db 153 GAGTTTATGAGAGGATGATGAGATCCGATGAGGATGAGAAATACCTCGTTGATGAA 212  
Qy 814 CGCGCAGACAGACTGAACTCCACCTGCTGCTCCCTGCTCCGATGATGCGGACATG 873  
Db 213 TATGCTACCAATCAACGTTTACCTTATTAATGATGATGCGTGGAGAGCGGATG 272  
Qy 874 CAGAGAACATTCAGACTTTTGTATGATGATGATGATGATGATGATGATGATGATG 933  
Db 273 CAAAGAGAGAGGTTGAGCTTAATGCAAAATGCTAAATTTCTGCTTTTATATATAT 332  
Qy 934 CGCGAGCGCATTTTCCGCTCTGCTGATTTTTCGCA---TCCACAGATGATGCGGCC 990  
Db 333 CAATCATCAAAATTTTCTATGATGATGATGATGATGATGATGATGATGATGATG 392  
Qy 991 ATGCTCTGAGGCTTCACTGCGCATCTCACTTTTGGGAAACCGGCGGCAATGATG 1050  
Db 393 ACAGCTAGTTTGGATGTTGGGGATATCACTTTTCCGAAACCGGATGATGATGATG 452  
Qy 1051 TTCCGAGGCTCTCGGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1110  
Db 453 TTGCAAGGTAAGAGTAATGAGAAACGTTGAAGATGAAAGTCCGAAAGGATACAA 512  
Qy 1111 CAGGCGGAGAAATTTGTAAGAACTGATGATGATGATGATGATGATGATGATGATG 1159  
Db 513 TCGGCTGAATTTTATTCGAAAGGCTGATTTGATGATGATGATGATGATGATGATG 561

RESULT 5  
B2548722 1063 bp DNA linear GSS 17-DEC-2002  
LOCUS B2548722  
DEFINITION pasci-60\_1377.s1 pasci-60 Pseudomonas aeruginosa genomic clone  
ACCESSION B2548722  
VERSION B2548722.1 GI:27152303  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 1063)  
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
Burns,J.L., Kaul,R. and Olsen,M.V.  
TITLE Whole-Genome-Sequence variation among multiple isolates of  
Pseudomonas aeruginosa library  
JOURNAL J. Bacteriol., (2002) In press  
COMMENT Contact: Chris K. Raymond

Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: Shotgun.

FEATURES  
source location/Qualifiers  
1..1063  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="1-60"  
/db\_xref="taxon:287"  
/clone="pasci-60\_1377"  
/note="clinical isolate 1-60 Whole genomic shotgun library."

BASE COUNT 202 a 316 c 335 g 209 t 1 others  
ORIGIN

Query Match 3.8%; Score 81.2; DB 29; Length 1063;  
Best Local Similarity 49.8%; Pred. No. 5.2e-12;  
Matches 206; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

Qy 1449 CACGCTTGGGCGGAGAGGCTGCGTTTGGCGGTCGTGGATTTGCTGGCGGAGCT 1508  
Db 58 CATCCCGCTCCGAGAGGCTATGCAAGGCTGCTGCTGATGAAATGCGCAAGCTT 117  
Qy 1509 AAACCTGCGATCTGTCATCATTCACACCTTCGCGCGCAATTTGCGAGCGGCTGA 1568  
Db 118 CAAGATCCGATCCTTCATTCATGACAGCCCGGCGCTTACCCGGGATGATGCGGA 177  
Qy 1569 GAGACTGCGATGCGAAGCTCGATGCGGCGACCTTCCAACTTATGACGCTCCCT 1628  
Db 178 GGAACGCGCCAAAGCGAGCGATGCTGAACTGCGGATATGCGGACTGAAGAC 237  
Qy 1629 CCCACCGTTTGGCTATTTATGTCAGGCGCTTGGCGCGCTGCGCCATGCTGCC 1688  
Db 238 GCGATATGCGACACCTGATGCGGAGGCGGCTTCCGGCGCGCTGCGCATCGGT 297  
Qy 1689 GCGGATCTGATCTTACGCGGCGGCAAAAGCGTGGCTGCGCATTTGCGACCAAGGCGC 1748  
Db 298 CTGCGACCAATGTAACATGCTGCAATCTCACCTATTCGATGATCGCGGAAGATG 357  
Qy 1749 CTGGCGCATCTCTTCCGCAACCAACCAACCGCGGGAATGATGAGGACAGGCGT 1808  
Db 358 CGCTCCATCTCTGGAAGACCGCGGAGAGCGCGGAAAGCCGCGAGGCAATGGGCT 417  
Qy 1809 GCAAGCGCACGCACTTTAAGCCAAAGGCTTATGACGAGGATGTCGCCGAAAC 1862  
Db 418 CACGCGCGAGCGGCTGAAAGCGCTGGGCACTGTCGCAAGGTCATGACGAGACC 471

RESULT 6  
B1310727 712 bp mRNA linear EST 20-JUN-2001  
LOCUS B1310727  
DEFINITION EST5132477 GESP Medicago truncatula cDNA clone pGESP8024 5' end,  
ACCESSION B1310727  
VERSION B1310727.1 GI:14985054  
KEYWORDS EST.  
SOURCE Medicago truncatula  
ORGANISM Medicago truncatula (barrel medic)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
Medicago.

REFERENCE 1 (bases 1 to 712)  
Grusak,M.A., Samac,D.A., Town,C.D., Van Aken,S., Utechtack,T., Cho  
J. and Fraser,C.M.  
TITLE ESTs from developing reproductive tissues of Medicago truncatula  
JOURNAL Unpublished  
COMMENT Contact: Michael A. Grusak

USDA/ARS Children's Nutrition Research Center  
Baylor College of Medicine  
1100 Bates Street, Houston, TX 77030-2600, USA  
Tel: 713-798-7044  
Fax: 713-798-7078  
Email: mgrusak@bcm.tmc.edu  
B397582e

TIGR sequence name: MTPAL607K  
More information is available at: www.medicago.org  
Seq primer: SKmod (CTA GAA CTA gtc gat cc).

## FEATURES

source

1..712  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/cultivar="A17"  
/db\_xref="taxon:3880"  
/clone="pGESD8j24"  
/tissue\_type="immature seeds"  
/dev\_stage="immature seeds, 11 to 19 days after  
pollination"  
/clone\_1lb="GESD"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Immature seeds, collected from pods ranging in age  
from 11 to 19 days after pollination, were harvested from  
greenhouse-grown plants. Seeds were removed and  
separated from pod walls and were immediately frozen in  
liquid nitrogen. Seeds throughout the age range were  
pooled for mRNA extraction. cDNA was prepared from polyA+  
enriched RNA. The cDNA was directionally ligated into  
the Uniap XR vector from Stratagene and packaged using  
Gigapack III Gold packaging extracts. Plasmids containing  
cDNA inserts were excised from the recombinant lambda-Zap  
phage using Ex-assist helper phage and propagated in  
XL01R cells."

BASE COUNT 222 a 107 c 153 g 230 t  
ORIGIN

Query Match 3.8%; Score 81; DB 12; Length 712;  
Best Local Similarity 48.4%; Pred. No. 4.8e-12;

Matches 256; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

634 AATCAAGCCTATGACAGACCTTGAGCGCGCTGCAAGCAGCCAAATGCGATGATCG 693  
63 AATGATGATTTCAAAATCGCTTGTATTTATCAAGACAGAACTGATTTACTGATGG 122  
634 GTAATTATGAGAGGACCTGAGGGGCAATCCGGTACCGCTATTTTGTCCATTTT 753  
123 GTTCAAGAGGACAGGCTCAAGTAATGATGCTTCTGAGCAATGATGATTAATGATTT 182  
754 TCTTCTCTGCGGGTTCTTGGGACAGGTCGCGTGGCATCATGAAAGCGATTCAC 813  
183 GAGTTTATGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 242  
814 CGCGCAGAGAGCTGAACTCCCACTGCTGCTCCCTGCTTCCGCTGCTGCGGCAAG 873  
243 TATGTTACCAATCAACGTTTACCTTATTAATGATGATGATGATGATGATGATGAT 302  
874 CAGAGAGCAATCGAGCTTTTCTCATGATGATGATGATGATGATGATGATGATGAT 933  
303 CAAGAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 362  
934 CGCGAGAGGAGCTTGGCTGCTGCTGATTTGCGCA---TCCAGATGATGATGATGATG 990  
363 CAATATCAATCAAAATGATTTATGATGATGATGATGATGATGATGATGATGATGAT 422  
991 ATGCGCTTCTGCGGGTTCACTGCGCATCTCACTTTTGGGAAACCGGCGCGCATAGT 1050  
423 ACAGCTAGTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 482  
1051 TTCTCGGCTCTCTGCGGGTGAAGTAACTGAGGCAATGCGCTTCCAGAGGCTGTGAG 1110  
483 TTGCGAGTAAAGAGTAAATGAGAAAGCTTGAAGATGAAATGCGGAGGATATACAA 542

Qy 1111 CAGCGGAGAAATTTGATGAAACTGATGATGATGATGATGATGATGATGATGATG 1159  
Db 543 TCGCGTGAATTTTATTTGAAAGAGGATGATGATGATGATGATGATGATGATGATG 591

RESULT 7  
CA990607  
LOCUS  
DEFINITION  
EST44115 GESD Medicago truncatula cDNA clone GESD-29C10, mRNA  
sequence.  
ACCESSION  
CA990607  
VERSION  
CA990607.1 GI:27523501  
KEYWORDS  
SOURCE  
ORGANISM  
Medicago truncatula (barrel medic)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; euroside 1; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
Medicago.  
1 (bases 1 to 580)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Unpublished  
Contact: Grusak, M.A.  
USDA/ARS Children's Nutrition Research Center  
Baylor College of Medicine  
1100 Bates Street, Houston, TX 77030-2600, USA  
Tel: 713 798 7044  
Fax: 713 798 7078  
Email: mgrusak@bcm.tmc.edu

TIGR sequence name: MTPBX17TK  
More information is available at: www.medicago.org  
Seq primer: SKmod (CTA GAA CTA gtc gat cc).  
Location/Qualifiers

## FEATURES

source

1..580  
/organism="Medicago truncatula"  
/mol\_type="mRNA"  
/cultivar="A17"  
/db\_xref="taxon:3880"  
/clone="GESD-29C10"  
/tissue\_type="immature seeds"  
/dev\_stage="immature seeds, 11 to 19 days after  
pollination"  
/clone\_1lb="GESD"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Immature seeds, collected from pods ranging in age  
from 11 to 19 days after pollination, were harvested from  
greenhouse-grown plants. Seeds were removed and  
separated from pod walls and were immediately frozen in  
liquid nitrogen. Seeds throughout the age range were  
pooled for mRNA extraction. cDNA was prepared from polyA+  
enriched RNA. The cDNA was directionally ligated into  
the Uniap XR vector from Stratagene and packaged using  
Gigapack III Gold packaging extracts. Plasmids containing  
cDNA inserts were excised from the recombinant lambda-Zap  
phage using Ex-assist helper phage and propagated in  
XL01R cells."

BASE COUNT 171 a 88 c 129 g 192 t  
ORIGIN

Query Match 3.8%; Score 79.8; DB 14; Length 580;  
Best Local Similarity 49.5%; Pred. No. 9.5e-12;

Matches 235; Conservative 0; Mismatches 237; Indels 3; Gaps 1;

688 GAATCGGTAATTTACTGAGAGGACCGTGGAGGCAATCCGGTACCGTATTTTCC 747  
2 GATCGGTTCAAAAGGACAGGCTCAAGTAATGATTTCTGAGCAATAGATATATG 61  
748 GATTTTCTCTCTGCGGGTTCTTGGGACAGGTCGCGTGGTCCCATCATGAAAGCG 807  
62 GATTTTGAATTTATGAGAGGATGATGATGATGATGATGATGATGATGATGATGATG 121

QY 808 ATTCACCGGCGACAGAGCTGAACCTCCACCTGCTGCTCCCTGCTTCGGGTGTCG 867  
 DB 122 ATGTAATATGCTACCAATCAACGTTTACCTCTTATATGATATGTCGCTGAGAGACG 181  
 QY 868 GCATGTCAGAGAAGCAATCGAGCTTTTGTTCATGATGTCGTCCATCAACCGGCTGTGACG 927  
 DB 182 CGTATGCAAGAGAAGATTTGAGCTTAATGCAAAATGGCTAAATTTCTGCTTTTAAAT 241  
 QY 928 CGTACCGCGAGGCGCATTTTGGCTTCTGCTGTAATTTGCCA---ATCCACAGATGGT 984  
 DB 242 AATTATCAATCAATCAAAAGTTATTTATGATGACGATATCTTACATCTCTACTACTGT 301  
 QY 985 GGGCGGCGGCTGCTGGGGTTCATCTGAGCATCTCACTTTGGGAAACCGGCGCGAG 1044  
 DB 302 GGGGTAACAGCTAGTTTGGAAATGTTGGGGATATCATCTTTCGGAACCGGATGCTTAC 361  
 QY 1045 ATAGGTTTCTGCTGCTGCTGCTGCTGAGTTAAACCTGAGGCGCATGCGCTTCAGACGCT 1104  
 DB 362 ATAGCATTTGCAAGTAAGATATTTGAGAAAGCTTGAAGATGGAAGTCCGCAAGT 421  
 QY 1105 GTGACGACGCGGAGAAATTTGGTGAACCTGCTGATTTGATGAAATTTGCTGC 1159  
 DB 422 ATACATCGGCTGAATTTTATTCGAAAAGGCTGATTTGATTCCTGTAACAC 476

RESULT 8  
 BG584921  
 LOCUS BG584921 792 bp mRNA linear EST 11-APR-2001  
 DEFINITION EST486683 MHAM Medicago truncatula/Glomus versiforme mixed EST  
 library cDNA clone PMHAM-20E21 5' end, mRNA sequence.  
 BG584921  
 ACCESSION BG584921 GI:13599985  
 KEYWORDS EST.  
 SOURCE Medicago truncatula/Glomus versiforme mixed EST library  
 ORGANISM Medicago truncatula/Glomus versiforme mixed EST library  
 Eukaryota; mixed EST libraries.  
 1 (bases 1 to 792)  
 AUTHORS Harrison,M.J., Liu,J., Town,C.D., Van Aken,S., Uteback,T., Cho,J.  
 and Fraser,C.M.  
 TITLE ESTs from roots of Medicago truncatula after colonization with  
 Glomus versiforme, 2001  
 JOURNAL Unpublished  
 COMMENT Contact: Harrison M.J.  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73401  
 Tel: 580-223-5810  
 Fax: 580-221-7380  
 Email: mjharrison@noble.org  
 Noble EST name: N380393e TIGR sequence name: MTDW357K More  
 information is available at: <http://www.medicago.org>  
 Seq primer: SKmod (CTA GAA CTA gfc gAT CC).  
 Location/Qualifiers

FEATURES  
 source  
 1..792  
 /organism="Medicago truncatula/Glomus versiforme mixed EST  
 library"  
 /mol\_type="mRNA"  
 /cultivar="Medicago truncatula genotype A17"  
 /db\_xref="taxon:119092"  
 /clone="PMHAM-20E21"  
 /tissue\_type="roots colonized with Glomus versiforme"  
 /dev\_stage="Roots harvested at 10, 17, 22, 31 and 38 days  
 post-inoculation with Glomus versiforme. The library was  
 made from a mixture of RNA from each of these stages."  
 /lab\_host="E. coli strain XL0LR"  
 /clone\_lib="MHAM"  
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:  
 XhoI; cDNA was prepared from polyA+ enriched RNA from  
 roots harvested at 10, 17, 22, 31 and 38 days  
 post-inoculation with Glomus versiforme. The cDNA was  
 directionally ligated into the Unizap XR vector from  
 Stratagene and packaged using Gigapack III Gold packaging

BASE COUNT 264 a 112 c 151 g 265 t  
 ORIGIN  
 Query Match 3.8%; Score 79.8; DB 10; Length 792;  
 Best Local Similarity 49.5%; Pred. No. 1.1e-11;  
 Matches 235; Conservative 0; Mismatches 237; Indels 3; Gaps 1;  
 Extracts, Plasmids containing cDNA inserts were excised  
 from the recombinant lambda-Zap phage using Ex-assist.  
 helper phage and propagated in XL0LR cells."

QY 688 GAATCGGTAATTAAGTGAAGAGCAGCCGAGAGGCAATCCGATGCGGCTTATTTGTC 747  
 DB 33 GATCGGTTAAACAGGACAGGTCAGTAAGTATTCCTGATGCAATAGGATATATG 92  
 QY 748 GATTTTCTTCTCTGCGCGTTCCTTTGGCAGCGTCCGTCGTCGATCAATGAAGCG 807  
 DB 93 GATTTTGAATTAAGGAGGATGATGGAATCCGATGAGGATGAAGAAATCACTCGTTG 152  
 QY 808 ATTCACCGGCGACAGAGCTGAACCTCCCACTGCTGCTCCCTGCTTCGGGTGTCG 867  
 DB 153 ATTCATATGCTACCAATCAACGTTTACCTTATTTATGATATGTCGCTGAGAGCG 212  
 QY 868 GCATGACAGAGAAGCAATCGAGCTTTTGTATGATGTCATTAACCGGCGCTGTCAG 927  
 DB 213 CGTATGCAAGAGAAGATTTGAGCTTATGCAATGCTAAATTTCTGCTTTTATAT 272  
 QY 928 CGTACCGCGAGCGCATTTTCCGCTTCTGCTGATTTGGCA---ATCCAGATGGT 984  
 DB 273 AATTATCAATCAATCAAAAGTTATTTATGATGACGATCTTACATCTCTACTACTGT 332  
 QY 985 GGGCGGCGGCTGCTGGGGTTCATCTGAGCATCTCACTTTTGGGAAACCGGCGCGAG 1044  
 DB 333 GGGGTAACAGCTAGTTTGGAAATGTTGGGGATATCATTTTCCGGAACCGATGCTTAC 392  
 QY 1045 ATAGGTTTCTGCTGCTGCTGCTGCTGAGTTAAACCTGAGTGAATTTGCTGC 1104  
 DB 393 ATACATTTGCAAGGTAAGATATTTGAAGAAACGTTGAAGATCGAAGTCCGAGGT 452  
 QY 1105 GTGACGACGCGGAGAAATTTGGTGAACCTGCTGATTTGATGAAATTTGCTGC 1159  
 DB 453 ATACATCGGCTGAATTTTATTCGAAAAGGCTGATTTGATTCACCTGTAACAC 507

RESULT 9  
 BZ561247  
 LOCUS BZ561247 1042 bp DNA linear GSS 17-DEC-2002  
 DEFINITION pacs2-164\_3167.y3 pacs2-164 Pseudomonas aeruginosa genomic clone  
 BZ561247  
 ACCESSION BZ561247  
 VERSION BZ561247  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 1 (bases 1 to 1042)  
 AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
 Burns,J.D., Kaul,R. and Olsen,M.V.  
 TITLE Whole-Genome-Sequence Variation among multiple isolates of  
 Pseudomonas aeruginosa library  
 JOURNAL J. Bacteriol., (2002) In press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: [craymond@u.washington.edu](mailto:craymond@u.washington.edu)  
 Class: shotgun.  
 Location/Qualifiers

FEATURES  
 source  
 1..1042  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"

/strain="2-164"  
/db xref="taxon:287"  
/clone="pac2-164.3167"  
/note="clinical isolate 2-164 whole genomic shotgun library."  
BASE COUNT 207 a 320 c 270 g 244 t 1 others

Query Match 3.7%; Score 78.8; DB 29; Length 1042;  
Best Local Similarity 50.3%; Pred. No. 2.6e-11;  
Matches 194; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

1477 GCGCTGCGGATTCGTCGCGCGCGAGCTGAACCTGCCATCTGTCATCATTCAGC 1536  
110 GCTGCGCGCTATGTAATGCGCAACCTTCAATGCGATCCCTCACTTCATCGAC 169  
1537 ACCTCGCGCGCGAATGTCGAGCGCGGTGAGAGCTCGGAGTCGCAAGCTGATTCG 1596  
170 ACGCCCGCGCGCTACCGCGGCGATCGATGCGGAGAACCGGCGCAGAGCGGATCGCC 229  
1597 GCGACCTTGTCCAGCTTATCGACGCTCCCTCCCACTGTTGCTCATTTGTCAG 1656  
230 TGGAACTGCGGCGGTATGCGCGCGACTGAAGACCGCGATCATCGCCACCTGATCGCGAG 289  
1657 GCGCTTGGCGGTGCGCGCGCGCTGCGCATGTCGCCCGCATCTGTCATGCGCGCGAAGC 1716  
290 GCGGCTTCCGCGCGCGCGCTGCGCATGCTGTCGCGACCATTTGAACATGCTGCAATTC 349  
1717 GCGTGCCTGTCGCGATTCGCGACGAGCGCGCTCGCGCATCTCTCCGCGACACCAAC 1776  
350 TCCACCTATTCGCGTATCTCGCGGAGGCTGCGCTCATCTCTCGAAGACCGCGAG 409  
1777 CACGCGCGGAAATCATAGAGGACAGCGCTGCGAGCGCGACCTTTAAGCCAGG 1836  
410 AAGGCGCGGAGACCGCGCGCGCATGCGCGCATCAACGCGCGCTGAAAGCGCTGGGCG 469  
1837 CTTATCGACGCGGATCGTCGCGAAGC 1862  
470 ATCTGCGACAGCTCATCGACCAACC 495

RESULT 10  
LOCUS B1725661 539 bp mRNA linear EST 19-SEP-2001  
DEFINITION 1031080C09.y1 C. reinhardtii CC-1690, Stress II (normalized),  
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.  
ACCESSION B1725661  
VERSION B1725661.1 GI:15701356  
KEYWORDS EST  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadaceae; Chlamydomonas.  
REFERENCE 1 (bases 1 to 539)  
AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre,  
P., McDermott, J. P., Shlager, U., Sillow, C. and Stern, D.  
TITLE Analysis of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants. Project: 1031  
JOURNAL Unpublished  
COMMENT Contact: Charles Hauser  
PCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu

FEATURES  
SOURCE location/Qualifiers  
1..539  
/organism="Chlamydomonas reinhardtii"  
/mol type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"

/db xref="taxon:3055"  
/clone.lib="C. reinhardtii CC-1690, Stress II (normalized)  
), Lambda Zap II"  
/note="vector: pBluescript II SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Stress condition II library, constructed by John  
Davies and Jeffrey McDermott, combines cDNAs from CC-1690  
cells grown to mid-log phase in TAP (NH4+ - containing)  
and shifted to TAP - NO3- (24hrs); H2 production  
conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant  
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +  
sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).  
PolyA mRNA was purified from each sample, pooled and cDNA  
synthesized. The cDNA was directionally cloned into Lambda  
Zap II (Stratagene) in the EcoRI (5') and XhoI (3')  
sites. pBluescript II SK- plasmids were excised from the  
lambda Zap clones by superinfection with ExBast  
(Stratagene) phage. The library was normalized using  
method 4 described in Bonaldo et al., (1996) Genome  
Research 6: 791-806."  
BASE COUNT 95 a 185 c 175 g 84 t  
ORIGIN

Query Match 3.7%; Score 78.4; DB 12; Length 539;  
Best Local Similarity 51.0%; Pred. No. 2.3e-11;  
Matches 210; Conservative 0; Mismatches 201; Indels 1; Gaps 1;

1449 CACGCTTGGCGCGCGAGAGCTGCTTGGCGGTGCGCATTCGTCGCGCGAGCT 1508  
56 CATGCCCGCGCAACGCGCTACCGCGAGCGCGCTGCTTACGCGCGCGCAACATTT 115  
1509 AAACCTGCGCATTCGCTCATCATCATGACACCTCGCGCGCGCAATGTCGCGCGCTGA 1568  
116 TGGTCTGCCATCATCATCTTCGTGACACCGCGGACCTTATGCGCGCAAGCGCGGA 175  
1569 GAGAGCTCGCGCATTCGCACTGATTCGCGCGACCTTGTCAAGCTTATCGAGCTCCCT 1628  
176 GGAAGCTGGCGCGAGCGCGCGCATTCGCGTGAAGAGTTCGCGCTGCGCTGCT 235  
1629 CCCACCGTTTGGTCAATTTATGTCAGGCGGTGCGCGCGCGCTGCGCGCATGCTGCC 1688  
236 GCCCATCATCTCGGCGGTGCTATGCGCGAGGCGCGCTGCGCGCGCGCGCATGCTGCC 295  
1689 GCGCATCTGCTTACCGCGCGCGCGCGCGCGCGCTGCTGCGCGCATTCGCGCGCGCGC 1748  
236 GCGCAACCGCAACCTGATCATGAGAGAGCGGCTTACTATCGCGCTGCGCGCGCGCGC 355  
1749 CTCGCGCATCTCTTTCGCGCA-CACCAACACAGCGCGCGGAAATCATAGAGCGACAGGCG 1807  
356 GCGCGCATCTCTGCGGAGAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 415  
1808 TCGAGGCGCGACGCACTTTAAGCCAGGAGCTTATCGAGCGGATGCTCGCGCA 1859  
416 TCACCTCGCGCGAGCTGATGAAGTTCGCGCGCGCATGAGCACCATGCTGTCGCGCA 467

RESULT 11  
LOCUS AG247247 652 bp DNA linear GSS 13-DEC-2002  
DEFINITION Locus japonicus DNA, clone: LJT20K03\_not, genomic survey sequence.  
ACCESSION AG247247  
VERSION AG247247.1 GI:26647012  
KEYWORDS GSS.  
SOURCE Locus japonicus  
ORGANISM Locus japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteeae;  
Locus.  
REFERENCE 1  
AUTHORS Sato, S., Nakamura, Y. and Tabata, S.  
TITLE Locus japonicus TAC End sequences  
JOURNAL Published only in Database (2002)  
REFERENCE 2 (bases 1 to 652)

AUTHORS Sato, S.  
TITLE Direct Submission  
JOURNAL Submitted (20-NOV-2002) Shusei Sato, Kazusa DNA Research Institute,  
The First Laboratory for Plant Gene Research; 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan  
(E-mail: sato@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/  
Tel: 81-438-52-3935 (ex. 2336), Fax: 81-438-52-3934)  
FEATURES  
source  
1. 652  
/organism="Lotus japonicus"  
/mol\_type="genomic DNA"  
/strain="Miyakojima MG-20"  
/db\_xref="taxon:34305"  
/clone="UJT20K03 not"  
/clone\_lib="genomic TAC library"  
/note="VECTOR: pYITAC7"  
BASE COUNT 198 a 107 c 129 g 218 t  
ORIGIN  
Query Match 3.7%; Score 77.6; DB 29; Length 652;  
Best Local Similarity 50.0%; Pred. No. 4.5e-11;  
Matches 222; Conservative 0; Mismatches 219; Indels 3; Gaps 1;  
QY 719 AGGGCATTCGGTAGCCGTTATTTTGTCCGATTTTCTCTCCGCGGTTCTTTGGGCA 778  
DB 72 ATGGATTCTGTAGCAATGGGATTATGATTTGATTTATGGAGTATGATGGAT 131  
QY 779 CGGTCCGTCGGTGGCATCATGAAGGCGATTCACCGCGCCACAGAGCTGAATCCGAC 838  
DB 132 CCGTAGTAGTGAGAAATCAACCCGTTGGTGTGATATGATCAACCAACTTTTACCTTC 191  
QY 839 TGCTGATCTCCCGTCTCCGGTGTGGCGCATGAGAAAGCAATGAGATTTTGTCA 898  
DB 192 TTATGTGTATGTATGATCTGAGAGACATGATGCAAGAGAGAGATTGATGATGATC 251  
QY 899 TGATGTGTCCATTAACCGCGGCTGTGACGCTCAACCGAGGCGCATTTGCCGTTCTCG 958  
DB 252 AAATGGCTAAATATCTTCTGTATATGATTAATCACTAAATTAAGTTATTTCTATG 311  
QY 959 TGTATTTTCCGCA--TCCACAGATGGTGGCGCATGCTCGTGGGTTTCAATGGGC 1015  
DB 312 TATCAATCTTAATCTCCCACTACTGTGGGTTAAAGTATTTTGGCATGTTGGGGG 371  
QY 1016 ATCTCACTTTTGGGAACCGCGCGCGCATAGTTTCTGGGTCTCGCGTGTGAGT 1075  
DB 372 ATATCATTAATGCGAACAATGCTTACATTTGATTTGCGGTAAGAGTAATGAC 431  
QY 1076 TAACCACTGGGATGCGCTTCAGACGGTGTGACAGACGAGAAATTTGTGAAGCTG 1135  
DB 432 AAACCTGAATTAAGCAATTCGCGAAGTTTCAAGAGGCGGCAAGATTTATTCATAGG 491  
QY 1136 GTGTGATGATGATGATTTGTGCG 1159  
DB 492 GCTTGTGATTAATCAATGTAAC 515  
RESULT 12  
CB977199/c 617 bp mRNA linear EST 01-MAY-2003  
LOCUS CB977199  
DEFINITION CAB40003\_1vA\_Ra\_D11 Cabernet Sauvignon Berry - CAB4 Vitis vinifera  
ACCESSION CB977199  
VERSION CB977199.1 GI:30300405  
KEYWORDS EST.  
SOURCE Vitis vinifera  
ORGANISM Vitis vinifera  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; Vitaceae; Vitis  
1 (bases 1 to 617)  
Goes da Silva, F., Iandolo, A., Lim, H., Baek, J., Jones, K. and Cook  
, D.  
Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'

JOURNAL Unpublished  
COMMENT Contract: Douglas Cook, PhD  
CABs Genome Facility  
UC Davis, Plant Pathology  
One Shields Ave, Davis, CA 95616, USA  
Tel: 530 754 6561  
Fax: 530 754 6517  
Email: drccook@ucdavis.edu  
Seq primer: GCCAATGATGCTAG.  
FEATURES  
source  
1. 617  
/organism="Vitis vinifera"  
/mol\_type="mRNA"  
/cultiivar="Cabernet Sauvignon"  
/db\_xref="taxon:29760"  
/clone="CAB40003\_1vA\_Ra\_D11"  
/sex="hermaphrodite"  
/dev\_stage="Berry on stage II, 9 mm"  
/lab\_host="DH5alpha"  
/clone\_lib="Cabernet Sauvignon Berry - CAB4"  
/note="Organ: Berry; Vector: pDNR; Site 1: Sfil; Site 2:  
Sfil; CAB4 is a cDNA library of Vitis vinifera cv.  
'Cabernet Sauvignon' Clone 8 berries. Sampled berries were  
collected from field-grown vines during stage II of berry  
growth (berries were green and hard) at approximately 60  
days after full bloom. The average berry size was 9  
millimeters. Sampled vines were located at the University  
of California, Davis, Experimental Vineyard. cDNAs were  
made by oligo-dT priming and directionally cloned. 5' and  
3' adaptors were used in cloning as follows:  
5'-AAGCAGGTATCAAGCAGAGTGGCATTAAGCGCGG-3' and  
5'-ATCTTAAGCCGAGCGGCGGCGGCAATG-dt(30)NN-3'. Library was  
constructed using the CloneTech Creator SMART kit and  
size-selected to contain the 0.5-3 kb size fraction."  
BASE COUNT 191 a 129 c 94 g 203 t  
ORIGIN  
Query Match 3.5%; Score 75; DB 14; Length 617;  
Best Local Similarity 47.8%; Pred. No. 2.5e-10;  
Matches 250; Conservative 0; Mismatches 270; Indels 3; Gaps 1;  
QY 640 GGTATCAGAGACCTTGAGCGGCTCGAAGCAGGCAATGCGATGCGTAT 699  
DB 616 GGTATTAAGAGCGTATGATTTTATCAAGAAAGCAGATTAAGTGGCTTCAA 557  
QY 700 ACTGAGAGGACCGGTGAGGCGATTCGGTACCGTTATTTTTCGATTTTCTTC 759  
DB 556 ACAGGACAGGTCAACTTAACGGATTTCCGTAGCAATGGGTTATGATTTCAAGTT 497  
QY 760 CTGGCGGTTCTTTGGCAGGTCGCGTGGTCCGATCATGAAGGATTCACCGGCGC 819  
DB 496 ATGGGGGTAGTATGGATCCGATAGTGAAGAAATTAACCGTTGATGCAATATGCC 437  
QY 820 ACAAGCTGAACCTCCCACTGCTGTCCTCCCTGCTCCGCGTGTGGCGCATGACAGAA 879  
DB 436 ACCAATGAATTTTCTACTCTTATTTTATGATGTTCTTCGAGAGACGATAGCAAGAA 377  
QY 880 GACAATGACCTTTTGTATGATGATGTCATTAACCGCGGCTGTGACGCTCAACCGGAG 939  
DB 376 GGAAGTTGAGCTTGAATGAGTGAATATCTTCGCTTATTAATGATTAATCAATCA 317  
QY 940 GCGATTTGGCGTTCCGCTGTATTTGGCA--TCCACAGATGGTGGCGGCATGCGC 996  
DB 316 AATTAAGATTAATTTATGATCAATCTTCAATCTTCACTAGTGGGTTGACAGCA 257  
QY 997 TCGTGGGCTCATCTGGGCACTCACTTTTGGGAAACCGGCGCGCATAGGTTTCTG 1056  
DB 256 AGTTTGTATGTTGGAGATATCATTTATGCGGAACCAATTCATCATTTGATTTGCG 197  
QY 1057 GGTCTCGCGTGTGAGTTAACCACTGGGCGATGCGTTTCAAGCGGTGTGACAGAGCG 1116  
DB 196 GTTAAAGGTTAATTGAACAACATTTGAAGAAAGACGATCTGAAGGTTACAGAGCGCT 137



Oy	1117 GAGAAATTGGTGAACACTGCTGTGATTAATGAATTGTGC	1159
Dd	136 GAAATATTATTCATAAGGCGTTAATTCGATTAAATCGAACAC	94
RESULT 13	A1229631	668 bp mRNA linear EST 20-JAN-1999
LOCUS	A1229631	
DEFINITION	EST226326 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone	
ACCESSION	REMCUJ63 3' end, mRNA sequence.	
VERSION	A1229631	
KEYWORDS	EST.	
SOURCE	Rattus sp.	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 668) Lee,N.H., Glodex,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D. Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat Gene Index Unpublished Contact: Lee, NH The Institute for Genomic Research 9712, Medical Center Drive, Rockville, MD 20850, USA Tel.: (301)-836-3529 Fax: (301)-838-0208 Email: nhlee@igf.org Seq primer: M13-21. Location/Qualifiers 1..668 /organism="Rattus sp." /mol_type="mRNA" /db_xref="ATCC (inhost):2036524" /db_xref="taxon:10118" /clone="REMCUJ63" /dev_stage="embryo 8, 12, 18 dpc" /clone_lib="Normalized rat embryo, Bento Soares" /note="Vector: pYT3pac; Site_1: EcoRI; Site_2: NotI"	
BASE COUNT	155 a 171 c 196 g 146 t	
ORIGIN		
Query Match	3.5% Score 74; DB 9; Length 668;	
Best Local Similarity	53.9%; Pred. No. 5.2e-10;	
Matches	152; Conservative 0; Mismatches 130; Indels 0; Gaps 0;	
Oy	707 AAGCACCGTGAGGAGCATTCGCGTAGCCGTTATTTGTCGATTTTTCTCCTCGAGC	766
Dd	129 AAGGACCTGTGAATGAAATGCGGGTGTGCGTGGGCAATTCGAGTTATGAGCG	188
Oy	767 GTTCTTTGGCAGCGTCCGCTGCGTGGGCCATCATGAAGCGCATCCGCGCCACAGC	826
Dd	189 GTTCATAGGGGTCTGTTGTGGGTGACCTTTCGTCGCGGCCGTGAGACAGCGCTGGAAG	248
Oy	827 TGAACCTCCCACTGTGTGCTCCCTGCTCCGATGCGGAGCGCGCATCAGAAACATC	886
Dd	249 ATTAATGCCCGCTGATCTGCTTCTTCGCTGTGGGCGACGATGACAGAGCACTGA	308
Oy	887 GAGCTTTGTGATGATGTGTTCATMAACCGCGGCTGTGCAAGCTCACCGAGCGCAT	946
Dd	309 TGTCGTCATGCAATGGCGAAAACCTCTGGCGGCACTGGCAAAAATGCAGAGAGCGCGCT	368
Oy	947 TGCGCTTCTGTGTATTTGGCGCAATCCCAAGATGGGTGGCG	988
Dd	369 TGCGTATCATCTCCGTGCTGACCAACCGAGATGGCGGGT	410
RESULT 14	BZ035752	677 bp DNA linear GSS 09-OCT-2002
LOCUS	BZ035752	

[illegible]





Db 301 GTGCAATAGTTAACGGGCTTCACAGTCACCATCTCTCCGGTTTATAGTTATGTTCCGG 360  
Qy 361 ACGTCTAGGCAAAAGTAGTTTGTGAGATGAAAGCAATCCGTCATTTTATTCGGA 420  
Db 361 ACGTCTAGGCAAAAGTAGTTTGTGAGATGAAAGCAATCCGTCATTTTATTCGGA 420  
Qy 421 TCGATAGCTTAAATTTGGCTTAGATCTTCCGCTCTAAATAGTATGAGAGACATTGCA 480  
Db 421 TCGATAGCTTAAATTTGGCTTAGATCTTCCGCTCTAAATAGTATGAGAGACATTGCA 480  
Qy 481 ATTAAATTAACAAGCCATTTTCCGCGCTGAGAAAGCTTTCCGATATGATGATGAGG 540  
Db 481 ATTAAATTAACAAGCCATTTTCCGCGCTGAGAAAGCTTTCCGATATGATGATGAGG 540  
Qy 541 ATGGAACAACCTTCAGCATTTGACGCTCATPAGCTGCTTGGATGAGCCCTGACAGCTTCA 600  
Db 541 ATGGAACAACCTTCAGCATTTGACGCTCATPAGCTGCTTGGATGAGCCCTGACAGCTTCA 600  
Qy 601 TCTTGAATGAAACTCCCAATATGACAACTCAATCAAGGCTATGACAGACCTTGGAG 660  
Db 601 TCTTGAATGAAACTCCCAATATGACAACTCAATCAAGGCTATGACAGACCTTGGAG 660  
Qy 661 CGGGCTCGAAGCAAGCCAAATGCGATGATGCTAATCTGAGAAAGCAAGCTGGAG 720  
Db 661 CGGGCTCGAAGCAAGCCAAATGCGATGATGCTAATCTGAGAAAGCAAGCTGGAG 720  
Qy 721 GGCATTCGGATAGCGGTATTTGTGCGATTTTCTCTCCGCGGCTTCTTGGGACG 780  
Db 721 GGCATTCGGATAGCGGTATTTGTGCGATTTTCTCTCCGCGGCTTCTTGGGACG 780  
Qy 781 GTCGCGTGGATGCGCATATGAAAGCGATTCACCGCGCAGACAGACTGAACTCCCACTG 840  
Db 781 GTCGCGTGGATGCGCATATGAAAGCGATTCACCGCGCAGACAGACTGAACTCCCACTG 840  
Qy 841 CTGCTCTCCCTGCTTCCGCTGCTGCTGCGCATGACGAAAGCAATCGACTTTTGTCA 900  
Db 841 CTGCTCTCCCTGCTTCCGCTGCTGCTGCGCATGACGAAAGCAATCGACTTTTGTCA 900  
Qy 901 ATGGTGCATTAACCGGCGCTGTGCAAGGTCACCGGAGGCGCATTTCCGCTCTGCTG 960  
Db 901 ATGGTGCATTAACCGGCGCTGTGCAAGGTCACCGGAGGCGCATTTCCGCTCTGCTG 960  
Qy 961 TATTTGCGCAATCCGCGATGAGGTCGCGCATGCGCTGCTGAGGCTTCACTTGGGATTC 1020  
Db 961 TATTTGCGCAATCCGCGATGAGGTCGCGCATGCGCTGCTGAGGCTTCACTTGGGATTC 1020  
Qy 1021 ACTTTTGGGAAACCGGCGCGGAGATAGGTTTCTGCTGCTCTGCTGCTGCTGCTG 1080  
Db 1021 ACTTTTGGGAAACCGGCGCGGAGATAGGTTTCTGCTGCTCTGCTGCTGCTGCTG 1080  
Qy 1081 ACTGGGCAATGCGCTTCAGACGCTGTGACGAGCGGAGAAATTTGGTGAATGCTGCTG 1140  
Db 1081 ACTGGGCAATGCGCTTCAGACGCTGTGACGAGCGGAGAAATTTGGTGAATGCTGCTG 1140  
Qy 1141 ATTGATGAAATTTGTCGCACTCCCAATTTGCTGCAAGCGGTCGCAAAAACCTCAAGGTT 1200  
Db 1141 ATTGATGAAATTTGTCGCACTCCCAATTTGCTGCAAGCGGTCGCAAAAACCTCAAGGTT 1200  
Qy 1201 ATTGATGAAATTTGTCGCACTCCCAATTTGCTGCAAGCGGTCGCAAAAACCTCAAGGTT 1260  
Db 1201 ATTGATGAAATTTGTCGCACTCCCAATTTGCTGCAAGCGGTCGCAAAAACCTCAAGGTT 1260  
Qy 1261 GTGATGAGAGCGATTTGCGCTTCTGTCACCGCGAGAGGCTTGGAAATGCGGAGATTA 1320  
Db 1261 GTGATGAGAGCGATTTGCGCTTCTGTCACCGCGAGAGGCTTGGAAATGCGGAGATTA 1320  
Qy 1321 GAAACGTTGGGCGGAGAGCTGCTCAAGCTTTCTGCTGCTGCTGCTGCTGCTGCTG 1380  
Db 1321 GAAACGTTGGGCGGAGAGCTGCTCAAGCTTTCTGCTGCTGCTGCTGCTGCTGCTG 1380  
Qy 1381 GCTGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440  
Db 1381 GCTGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440

Qy 1441 GCGCGCTTCAAGCTTTGGGCGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500  
Db 1441 GCGCGCTTCAAGCTTTGGGCGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500  
Qy 1501 GCGAGCTTAAACCTGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560  
Db 1501 GCGAGCTTAAACCTGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560  
Qy 1561 GCGGCTGAGAGAGCTGCGCATGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620  
Db 1561 GCGGCTGAGAGAGCTGCGCATGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620  
Qy 1621 GCTCCCTCCCAACCGCTTGTGCTATTAATGTCAGGAGCTTGGCGGCTGCGGCTGCG 1680  
Db 1621 GCTCCCTCCCAACCGCTTGTGCTATTAATGTCAGGAGCTTGGCGGCTGCGGCTGCG 1680  
Qy 1681 ATGCTGCGCGCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740  
Db 1681 ATGCTGCGCGCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740  
Qy 1741 GAGGCGCTGCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800  
Db 1741 GAGGCGCTGCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800  
Qy 1801 CAAGGCTGAGAGGCGACGCACTTTTAAAGCCAAAGGCTTAAAGCCGATGCTGCGGAA 1860  
Db 1801 CAAGGCTGAGAGGCGACGCACTTTTAAAGCCAAAGGCTTAAAGCCGATGCTGCGGAA 1860  
Qy 1861 ACCGAGCACTTTGTTGAAGAAATTTCTGCGCAATGCAATGCAATGCAATGCAATG 1920  
Db 1861 ACCGAGCACTTTGTTGAAGAAATTTCTGCGCAATGCAATGCAATGCAATGCAATG 1920  
Qy 1921 AACCAATCCGAGAGGCGGAGCGGACAGCTGCTTCAACAGATTTAGAGCTTAAAGC 1980  
Db 1921 AACCAATCCGAGAGGCGGAGCGGACAGCTGCTTCAACAGATTTAGAGCTTAAAGC 1980  
Qy 1981 TAAAGAAATTTATGCGCTGATCAATGATGATGAACACAGGAGTACGCGAGCAGTGG 2040  
Db 1981 TAAAGAAATTTATGCGCTGATCAATGATGATGAACACAGGAGTACGCGAGCAGTGG 2040  
Qy 2041 GTGGCGGAAACCTTCAGAGGCGCTGTAAGCAGCTTGGCGGAATGTCAGTGAAGCTTC 2100  
Db 2041 GTGGCGGAAACCTTCAGAGGCGCTGTAAGCAGCTTGGCGGAATGTCAGTGAAGCTTC 2100  
Qy 2101 GCCGACCTTATGCTTGAATTC 2123  
Db 2101 GCCGACCTTATGCTTGAATTC 2123

RESULT 2  
US-09-362-899-2  
Sequence 2, Application US/09362899  
Patent No. 6361986  
GENERAL INFORMATION:  
APPLICANT: Degussa-His AG  
TITLE OF INVENTION: Prozess zur Herstellung von L-AMINO SÄUREN  
CURRENT FILING DATE: 1999-07-29  
EARLIER APPLICATION NUMBER: DE 19924365.4  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 2  
LENGTH: 1473  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1473)  
OTHER INFORMATION: accDA  
US-09-362-899-2



Dh 1008028: CGGTGACAGACCGGTGAGGAGACCGGTAATTGCGGCGCGGCGGTGAGCTGTGAGT 1007969  
Qy 752 TTTCTCTCTCGGCGGTTCTTTGGGACAGGTGCGGTGCGGATCATGAAAGCGATTC 811  
Db 1007968 TCGACTTCTGGGGCGGCTCGATTGGGGGAGGACGCGGCGAAATCGATCCGCGCGGTG 1007909  
Qy 812 ACCGCGGACAGAGCTGAAATCCCACTGTGTCTCCCTGCTTCCGAGTGTGCGGCA 871  
Db 1007908 AGCGGGGACCGCGGCGGCTGCGCTACGTGGCTACCAAGCTCGGAGAGGACCGCA 1007849  
Qy 872 TGCAGAAAGACATCGACTTTTGTGATGTGTTCATACCGCGGCTGTGACGCTC 931  
Db 1007848 TGCAGAAAGACAGGCTCGCTTTCAGATGTGTAAGATGCTGCGGCGCATTCAGCTGC 1007789  
Qy 932 ACCGCGAGGCGGCTTTGCGGTCTGTGTGTATTTGGCGAATCCCAAGATGGTGGCGCA 991  
Db 1007788 ACAACGAGCGCGCTGCGCTACCGGTCTATTGGCGCATCCGACCGAGGTGAGTTT 1007729  
Qy 992 TGGCTCGTGGGGTTTCATCTGAGGATCTCACTTTTGGCGAAACCGGCGCGGAGATAGGT 1051  
Db 1007728 TCGGCTGTGGGGCTCGCTGGGGGATCTCACCGTGGCGAGCGGCGCGCTGTAGCTGCT 1007669  
Qy 1052 TCTGGGCTCTGCGGTGTGAGATTAACTGAGCATGCGCTTCAGACGCTGTGACG 1111  
Db 1007668 TTCTGGGACCAACGGGTCTATGAGTTGCTCTATGGCGACCCCTTCCATCCGCGGTCCAAA 1007609  
Qy 1112 AGCGCGAATTTTGTGAAAATGCTGTGATTTGATGAAATGTGTGCACTCCCAATTGC 1171  
Db 1007608 CCGCGGAATCTACGCGGAGATGAGATCATCGAGGCGGTGTGCACTGAGACGCGCTAC 1007549  
Qy 1172 GTGACGCGGTGCAAAAACCTCAAGTTTATTACCGCGGTAGAGGACAGATCGTTT 1231  
Db 1007548 GACCGATCTGATGTGTGCTGTGAGCTGTGATGAGAGCTCCGAAACGCTTCGCGAC 1007489  
Qy 1232 CTCAACAATCTCTGCGGTGCA---CTTCCGTGATGAGGAGCATTCGCGTCTGTG 1288  
Db 1007488 CGCAGACCGCGCGCGCTGACCGGATGAGCCAGCTGGGACCTCGGTGTGCGATCGCC 1007429  
Qy 1289 ACCCGCAAGGCTTGGAAATCGGGAGATTATGAAACGTTGGGGGCAACGTCGTCAAGC 1348  
Db 1007428 GCGCGACCGCGCGGCGCTGACGAGCTACGACACGCGGCCACCAACCGGGGTGT 1007369  
Qy 1349 TTTCTGTGCGGCTCTGAGCGCATTTGAGCCGCGCTGTGCGCTTCCCTGGCGCGCATCG 1408  
Db 1007368 TGTGAGAAACCGATCAAGCGCAAGG---GCGACACGCTGTGTGCGCTGCGCTTTC 1007312  
Qy 1409 GCGGCGCGCGGTGTGCTGATTGGGACG-----GATGCGGCTTTCAGCGC 1453  
Db 1007311 GCGGCGCAACCAACGCTGTGCTCGGCGCAAAAGGCGAGTAGCGGCGGGGAAACATG 1007252  
Qy 1454 TTTGGGCGGAGAGCTGTGCTTTTGGCGCTGTGTGCAATTTCGTGGCGCGCGCAAGCTTAAAC 1513  
Db 1007251 TCGGCGCGCGCTGTGCTTACGCGAAAGCGACGCGGATGAGCGTTCGCGCGAGCTGTGCTC 1007192  
Qy 1514 TGCATGCTGTTCATTCATTCGACACTCGGCGCGAAATTGTGCGAGCGGCTGAGAGC 1573  
Db 1007191 TGCCTGTGTGTGCTCATTTAGCGGCGCGACCGCGCTTGTGCGCGGACCGAAGAG 1007132  
Qy 1574 TCGGATGCGCAAGCTGATTGCGCGCACCTTGTCCAGCTTATTCAGCGCTCCCTCCCA 1633  
Db 1007131 GCGGCGTGGCGCGGACAGATCGGCAATGCTGTGCGAGCTGCTCAAGCTGATACCCGA 1007072  
Qy 1634 CCGTTTGTGCTATTATTTGTGAGGCGGCTTGGGCTGTGGCGCTGTGCGCTGTGCGCGG 1693  
Db 1007071 CCGTGTGATCTGTGTGGCGGAGGCGAGCGGCGCGCGCTGTGCGATGTTGCGCGCG 1007012  
Qy 1694 ATTCGCTAGCGGCGCGAAACGCGGTGCTTCGCAATTGCGACAGAGGCGGCTCGG 1753  
Db 1007011 ACCGGGTGTGGCGCGGATCTCAAGGCTGTGCGCTTGTGCTTCCGAGAGGACAGCG 1006952  
Qy 1754 CCATCTCTTTCGCGACACCAACGAGCGCGGAAATCATAGAGGACAGAGCGCTGACG 1813  
Db 1006951 CGATCTGTTCGAGACACTGTCTCATGCGCGGAACTGTGCTGCGCGCAAGCATCCGGT 1006892

Qy 1814 CGACGCACTTTTAAGCCAAAGGCTTATGACGCGGATGCTGCGCGAAACCGAGCACTTTC 1873  
Db 1006891 CGGCGGACTTACTGAAAGTGGGGATTGTGACACCATCTGTCGGAGTACCCGACGCG 1006832  
Qy 1874 TTGAGGA 1880  
Db 1006831 CAGACGA 1006825

RESULT 4  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT FILING DATE: US/09/103.840A  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 14.7%; Score 312.6; DB 3; Length 4411529;  
Best Local Similarity 54.0%; Pred. No. 2.7e-80;  
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

Qy 572 ACTGCTTTTGAACCTTGACAGCTTCTTCTTGAATGAAATCCCAATATGACAAC 631  
Db 1008147 AGCGGTGCTAGACCGGGGATCTTTCGTACGTGGATATGACAGACCGCTGCGGTGCGG 1008088  
Qy 632 TCAATCAAGCTATGACAGACCTTGTGAGCGGCTGCAAGCAAGCCAAATGCGATGAAT 691  
Db 1008087 TAGCGCACTCTTATGCGCGGAGGTGCGCGCTCGGGCGGACCGGACCAAT 1008028  
Qy 692 CGTAAATTAATGAGAAAGGACCGGTGAGGCGATTCGGTATGCTTATTTTGTCCATT 751  
Db 1008027 CGGTGACAGCCGCTGAGGAGACGCTATTCTGAGCGGCGGCTGTGCTGTGAGT 1007968  
Qy 752 TTTCTTCTCGGCGGTTCTTTTGGGACGCGGTGCGGTGCGGATCATGAAAGCGATTC 811  
Db 1007967 TCGACTTCTGGGGGCTCGATTGTGAGTGTGAGAGCGCGCAACGATACCGCGCGCTG 1007908  
Qy 812 ACCGCGACAGAGCTAATACTCCCACTGTGTCTCCCTGCTTCCGATGTGCGGCA 871  
Db 1007907 AGCGGGGACCGCGGAGCGGCTGCGCTACTGTGCTACCAAGCTGGGAGGACCGGCA 1007848  
Qy 872 TGCAGAAAGACATCGAGCTTTTGTCAATGATGTTCATTAACCGGCTGTGAGCGCTC 931  
Db 1007847 TGCAGAAAGACAGCTGTGCTTCTGCAATGTGAAAGATGCTGCGGCAATCAAGCTGC 1007788  
Qy 932 ACCGCGAGGCGCATTTGCGCTGCTGTGATTTTGGCAATCCACAGATGGGTGGCGCA 991  
Db 1007787 ACAACGAGGCGCGCTGCTTACTGATGTATTGCGCATCCGACAGCGGATGGAATTT 1007728  
Qy 992 TGGCTGTGGGGTTTCATCTGAGCATCTCACTTTTGGCGAAACCGGCGCGGAGATAGGT 1051  
Db 1007727 TCGGCTGTGGGGCTCGCTGGGGATCTCAACCGTGTGCGAGCGGCGGCTGTATGCGCT 1007668  
Qy 1052 TCTGGGCTCTGCGGTGTGAGTTAACTGAGGCAATGCGCTTTCAGAGGCTGTGACG 1111  
Db 1007667 TTCTGGGACCAACGGGTCTATGAGTTGCTATGAGCGACCCCTTCCATCCGCGCTCCAAA 1007608



QY	1112	AGCGCGAAGATTGTGTAACCTGGTGTGATTGAGAAATTGTGCGACTCCAAATTGC	1171
Db	1007607	CCGCCGAAGATTACGGCGGACATGGGAATCATGACGGCGTGTTCATGTGAACCGCGTAAG	1007548
QY	1172	GTGCAACGGTGGCAAAAAACCTCAAGTATTACCCGGTAGAGCAACGATCGTTTTT	1231
Db	1007547	GACCGATGCTGGATGTGTGGTTGAGGGTGTCTATGACAGCTCCCGAACCCTTCCGGAC	1007488
QY	1232	CTTCAACAATCTCTGGCGTGCA---CTTCCGATGATGAGGCGATTGCGGTTCTGTG	1288
Db	1007487	CGCAACCCCGCGCCCGCTACCCGATGTGCCACCTGGGACCTCGTGTGTGCATCGCCGC	1007428
QY	1289	ACCCGCAAGGCGCTGGAATCGGGAGATTATGAAACGTTGGGGCAACGTCGCAAC	1348
Db	1007427	GCGCGAACCAGCCGGCGCTCAGGACGTACTACGACAGCGCCACCGACCGGGGTGTGT	1007368
QY	1349	TTTCTGTGTCGCGTGTGCGCGATTGACCCCGGCTGTGCGGATTGCCCTTGGCGGCAATG	1408
Db	1007367	TGTCAAGAAACCGATTCAAGGCGCAAGCG---GCAACACGCTGTGGCGCTTGGCCCGCTTGTG	1007311
QY	1409	GGGCGCGGCGCGTGTGCTGATTGGGCG-----GATGCGCGTTTCAAGC	1453
Db	1007310	GCGGCAACCCACGCTGTCTCTCGCCACGAAAGGCGAGTAGCGGCGGGGAAAGCACTG	1007251
QY	1454	TTGGGCGCGAAGAGCTGTGTTTGGCGCTCGTGSCATTTCGCTGGCGCGCGAGCTAAAC	1513
Db	1007250	TGCGGCGCGCTGTGCTTACGCGAAGCCGACCGGGATGGCGCTGCGCGCAAGCTGTGCC	1007191
QY	1514	TGCGGATGTGTTCATTCATCGAACCTTCGCGCGCGGAATTGTGCGAGGCGGCTGAGAGC	1573
Db	1007190	TGCGGCTGTGTCTGGTCAATTGACGCGGCGGAGCCGCGGTGTGCGGCGCGAACCGAAGG	1007131
QY	1574	TGGGATGTGCAAGCTCGATTGCGGGGACCTTGTCCAAAGCTTATGACGCTCCCTCCCA	1633
Db	1007130	GCGGCGTGGCGCGCAGATCGGCAATGCTCTGGCGCAACTGTCAACGCTGAGTACCCCGA	1007071
QY	1634	CCGTTTTCGATCATTTATGTCAGGGCGTTTGGCGTGGCGCGCTGAGCAATGTCGCCCG	1693
Db	1007070	CCGTGTGTGATCTGTGTGGCGCAAGGCGAGCGGCGGCGCGCTGGCATATGTGCTCCCG	1007011
QY	1694	ATCTGTGTTCACGCGCGCGAAAAACGCGTGTGTGTGCGATTGCAACAGAGGCGCTTCGG	1753
Db	1007010	ACCGGATGTGTGGCGCATCCACGCGCTGGCTGGCGCCCTTTCCTCCCGAGAGAACCGAGCG	1006951
QY	1754	CCATCTCTTTCGCGGCAACCAACACGCGCGGGAATATATGAGAGCAAGAGCGGTGAGG	1813
Db	1006950	CGATCTGTTCGAGACCATCTCTCATGTGCGCGAACTGTGCTGCGCCCAAGGCAATCCGGT	1006891
QY	1814	CGACAGCACTTTTAAGCCAGGCGCTTATCGACGAGATCTGTGCGCGAAACGAGACCTTTG	1873
Db	1006890	CGGCGGACCTTACTGAAGTGGGGAATTGTGCAACATCTGTGCGGAGTACCCCGACGCG	1006831
QY	1874	TTGAGAGA 1880	
Db	1006830	CAGACGA 1006824	

RESULT 5  
 US-08-311-731A-140  
 ; Sequence 140, Application US/08311731A  
 ; Patent No. 6583266  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SMITH, DOUGLAS  
 ; APPLICANT: MAO, JEN-I  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
 ; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR  
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 411  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
 ; STREET: 600 ATLANTIC AVENUE  
 ; CITY: BOSTON  
 ; STATE: MASSACHUSETTS

```

: COUNTRY: USA
: ZIP: 02210
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/311,731A
:   FILING DATE:
:   CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
:   NAME: GATES, EDWARD R.
:   REGISTRATION NUMBER: 31,616
:   REFERENCE/DOCKET NUMBER: C0044/7125
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 617/720-3500
:   TELEFAX: 617/720-2441
: INFORMATION FOR SEQ ID NO: 140:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 36063 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double
:   TOPOLOGY: circular
: MOLECULE TYPE: DNA (genomic)
:   HYPOTHETICAL: NO
:   ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Mycobacterium leprae
:
: US-08-311-731A-140

```

Query Match	6.7%;	Score 142;	DB 4;	Length 36063;
Best Local Similarity	57.1%;	Pred. No. 4.8e-31;		
Matches 296;	Conservative 0;	Mismatches 220;	Indels 2;	Gaps 2

QY	685	GATGAATCCGGTAAATTTCGTGAGAAAGGCAACCGAGAGGGCAATTCGCGTACCCGTAAATTTTG	744
Db	31425	GACGAATCGGTGCTGACACAGCCAGGGGGTGGTAATTCGGGCAACGGGTACGATTAATTTGTC	31486
QY	745	TCCGAATTTTTCTTCTCTGGCGGGTTCTTTGGGCACGGTGCGTGGTGGCATCATGAAG	804
Db	31485	TGCAAAATTCAGTTCCTCGGTGGCCGTATCGGGGTGGCAGCCGGAGGCAATCACCCGC	31544
QY	805	GGCAATTCACCGGCGCCACAGAGCTGAACTCCACATGCTGGTCTCCCTGCTCCGCTGGT	864
Db	31545	GCCGTGACGG-CCGAAGCTGAGCGGCTGGCCG-CTGTCCTCGCAAGTTCCATGGCC	31602
QY	865	GGCGCATGCGAGGAACAATCGAGCTTTTGTATGATGATGTGTCATTAACCGCGCTGG	924
Db	31603	ACCGTATGAGAGGAAGCAACCGTGGCTTCTGCAAGATGTGAAGATCGCTGGCGCTGC	31662
QY	925	CACGCTCACCGGAGGGCGCATTTGCCGTTCTCGTGTATTTGCGCATATCCACGATGGT	984
Db	31663	AAACTCTCAAAACGGGCGGGCTGCGCCCTTACTTAAGGGTAATTTGCTGTAATCCAGCACCGGC	31722
QY	985	GGCGCCATGGCTCTGAGGGGTTCACTCGGCATCTCACTTTGCGGAACCCGCGCGCAG	1044
Db	31723	GGGGTCTTGCGCGTGGGGGCTCGCTCGGCGCAATGTAACCGTGGCGGAGCGGGTGTCTTA	31782
QY	1045	ATAGTTTTCTGGGTCTCGCGCTGGTGGAGATTAAACATCGGCGCATGCGCTTCCAAACGCT	1104
Db	31783	ATCGGCTCTCTTGGGCTCTCGGGTGTATGAGTTGCTGTATGGGGAACCTTTCCGCTCCAC	31842
QY	1105	GTGCGACAGGCGGAGAAATTTGGTGAAAATCGGTGTGATTAATGGAATTTGTGCGCACATC	1164
Db	31843	ATCCAAACAGCGCGAAGATCTGCACCGCATGGGGTAAATGACGCGTATGTCACGCTGAGC	31902
QY	1165	CAATTTCGTGACAGCGGTGGCAAAAACCTCTCAAGCTTAT	1202
Db	31903	GGACTGCAACTGACCTTGAATCGTGGCTGACGATGAT	31940

## RESULT 6

```
US-09-252-991A-12180
; Sequence 12180, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12180
; LENGTH: 1020
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12180

Query Match      5.5%; Score 116.4; DB 4; Length 1020;
Best Local Similarity 52.0%; Pred. No. 2.7e-24;
Matches 261; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 686 ATGATCGGTATTACTGGAAGAAGCACCCTGAGGCGATTCCGGTAGCCCTATTATTGT 745
DB 473 AGGAGCGCGTATGTCATGACGCGGAGCTGCAAGGATGTCGGATGCTGCGCTGCGCT 532
QY 746 CCGATTTTCTCTCTCGCGCGGTTCTTGGGACAGCGTCCGTCGTCGTCATCATGAAG 805
DB 533 TCGAGTTCTCTTATGAGCGGTTGATGAGCGCATGTCGGCAGAGCTTCGTTCCG 592
QY 806 CGATTCAACCGGCGCAAGAGCTGAACTCCACTGCTGCTCTCCCTGCTTCCGCTGTG 865
DB 533 CAGCCAAAGTCGCTGAGAGCGCTGCGCTGATCTCTTCGCTCCGCGCGCG 652
QY 866 CCGCATGAGAGAAACATCGAGCTTTTGTATGATGTCCTCAACCCGGGCGTGGC 925
DB 653 CCGCATGAGAGAAACCTGATCTGCTGATGCAAGTGGCAAGACCTGGGCGTCTGG 712
QY 926 AGCGTCACCGGCGAGCGCATTTGCGTTCCTGATATTGGCAATCCACGATGGTG 985
DB 713 CCGCGCTCGCGAAGAGCATCCGTTCTGCTCGATTTGACGACCCGCTCTACGGG 772
QY 986 GCGCATGCGCTCTGCGGCTTATCTGGGCACTCACTTTGCGGAACCCGGCGCGAGA 1045
DB 773 GCGTTCCGCGACGCTGGGATGCTCGGCGACGTATCGTCCGCAACCCAGGCGCTGA 832
QY 1046 TAGGTTTCTGCGTCTCGCGTGGTGAAGTTAACACATGGGCGATCGGCTTCCAGACGGTG 1105
DB 833 TCGGCTTGGCGCGCTCCCGGATGACAGACGCTCCGCGAAGAGCTCGCGAAGGCT 892
QY 1106 TGCACAGGCGGAGATTTGTGAACACTGTGTGATGATGAATTGTGCGCACTCC 1165
DB 893 TCCAGCGTAGAGATTCTCTTGAGCATGGGCGCATGACATGATCGATCGGCGG 952
QY 1166 AATTGCGTCAAGCGGTGCAAA 1187
DB 953 AGTTGCGGCGCGCGCTGGCCAA 974

RESULT 7
US-09-252-991A-12415/c
; Sequence 12415, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
```

```

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12415
; LENGTH: 1338
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12415

Query Match      5.5%; Score 116.4; DB 4; Length 1338;
Best Local Similarity 52.0%; Pred. No. 3.1e-24;
Matches 261; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

QY 686 ATGATCGGTATTACTGGAAGAAGCACCCTGAGGCGATTCCGGTAGCCCTATTATTGT 745
DB 677 AGGAGCGCGTATGTCATGACGCGGAGCTGCAAGGATGTCGGATGCTGCGCTGCGCT 618
QY 746 CCGATTTTCTCTCTCGCGCGGTTCTTGGGACAGCTGCGTCCGTCGTCGTCATCATGAAG 805
DB 617 TCGAGTTCTCTTATGAGCGGTTGATGAGCGCATGTCGGCAGAGCGCTTCGTTCCG 558
QY 806 CGATTCAACCGGCGCAAGAGCTGAACTCCACTGCTGATCTCCCTGCTTCCGCTGTG 865
DB 557 CAGCCAAAGTCGCTGAGAGCGCTGCGCGATGATCTGCTCCGCTCCGCGCGGCG 498
QY 866 CCGCATGAGAGAAACATCGAGCTTTTGTATGATGTCCTATATGATGATGATGATGATG 925
DB 497 CCGCATGAGAGAAACCTGATCTGCTGATGCAAGTGGCAAGACCTCGCGGCTCTGG 438
QY 926 AGCGTCACCGGCGAGCGCATTTGCGTTCCTGATATTGGCAATCCACGATGGTG 985
DB 437 CCGCGCTCGCGAAGAGCATCCGTTCTGCTCGATTTGACGACCCGCTCTACGGG 378
QY 986 GCGCATGCGCTCTGCGGCTTATCTGGGCACTCACTTTGCGGAACCCGGCGCGAGA 1045
DB 377 GCGTTCCGCGACGCTGGCGATGCTCGGCGACGTATGTCGAGAGAGCCAAAGCGCTGA 318
QY 1046 TAGGTTTCTGCGTCTCGCGTGGTGAAGTTAACACATGGGCGATCGGCTTCCAGACGGTG 1105
DB 317 TCGGCTTCCGCGGCTCCCGGATGACAGACCGTCCGCAAGAGCTCGCGAAGGCT 258
QY 1106 TGCACAGGCGGAGATTTGTGAACACTGTGTGATGATGAATTGTGCGCACTCC 1165
DB 257 TCCAGCGTAGAGATTCTCTTGAGCATGGGCGCATGACATGATGATGATGATGATG 198
QY 1166 AATTGCGTCAAGCGGTGCAAA 1187
DB 197 AGTTGCGGCGCGCGCTGGCCAA 176

RESULT 8
US-09-773-816-1/c
; Sequence 1, Application US/09773816
; Patent No. 6340774
; GENERAL INFORMATION:
; APPLICANT: Stanford University
; APPLICANT: Khosla, Chaitan
; TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR
; FILE REFERENCE: 26600-20210.00
; CURRENT APPLICATION NUMBER: US/09/773,816
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/243,458
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/179,305
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 23673
; TYPE: DNA
```

ORGANISM: Human  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (1) (23623)  
 OTHER INFORMATION: n = A,T,C or G  
 US-09-773-816-1

Query Match 5.1%; Score 108.6; DB 4; Length 23673;  
 Best Local Similarity 46.6%; Pred. No. 2,7e-21;  
 Matches 611; Conservative 0; Mismatches 619; Indels 81; Gaps 5;

QY 684 CGATGATGCTGATTAATTAAGGAGGACCGTGAAGGGGCAATTCGGTACCGCTTAATTTT 743  
 DB 8617 CGACGAGGCGCTGCTGATGTCACCGCGGATCCACGCGCAGCCGCTGTCGTCGCCG 8558  
 QY 744 GTCCGATTTTCTCTGCGGCTTCTTTGGCAGGTCGCTCGGTGCGATATGA 803  
 DB 8557 GATGACCTTCGCTTCTGCGGCGGACCTGCGGCGGCGGTGCGGAATGATCACTCT 8498  
 QY 804 GCGCATTCACCGCGCCACAGAGCTGAATCCCACTGCTGTCTCCCTGCTTCCGGTGC 863  
 DB 8497 CGCCCGCGAGACCGCGCTCGCCGAACACTGCGCTGTGATGTACGCGCTTCGGCGG 8438  
 QY 864 TCCGCGCATGCAAGAAACATGAGCTTTTGTATGATGTGTCCATAACCGCGCTGT 923  
 DB 8437 CGCGGCGATGAGAAAGCGCGCTGCGCTCATGAGATGGCCAGACAGCGCGCCCT 8378  
 QY 924 GAGGCTACCGCGGCGCCGATTTGCGCTTCTGCTGTATTTGCGCAATCCACGATGG 983  
 DB 8377 CGGGCACTGAGAGAGCGCGCGGTGACATCTGCGTGCACGACCCCACTTAACG 8318  
 QY 984 TGGCCCATGCGCTCGTGGGGTTCACTCTGGGATCTCACTTTTGGGAAACCGGCGCGCA 1043  
 DB 8317 CGGCGTCCGGGCGCTCTTGGCACTCTGTGCACTGTCGCGGAAACCGGCGCGCG 8258  
 QY 1044 GATAGCTTCTGCTGCTCTGCTGTGAGTTAACTGAGGATGCGCTTCCAGACG 1103  
 DB 8257 TCTCGGCTTCCGCGCGCGGCTGATCCGCGAGACATCCGAGAGAACTGCGCGCGCA 8198  
 QY 1104 TGTGAGAGCGGAGAAATTTGTGAAAATGCTGTGATGATGATGATGCTGCGCACT 1163  
 DB 8197 GTTCAAGACCGCGCAATACCTCTCGAACAACGCGCTGATGACCTGATGTCGCGCGCG 8138  
 QY 1164 CCAATTGCGTGCAGCGGTG-----CAAAACCC 1192  
 DB 8137 CGCGCTCGCGCGCACTGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 8078  
 QY 1193 TCAAGGTATTCAGCCGCTGAAGGCAACGAT-----CGTTTTTCTCCAA 1237  
 DB 8077 GAGCGGGGAACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8018  
 QY 1238 CAATCTCTGCGCTGCACTTCCGCTGTGAGAGCGGATTCGCGCTTCTGTAACCGCGCA 1297  
 DB 8017 CCCCCTCCGACCTCGCGCGCGCAAGACCCGCGCGCGCGCGCGCGCGCGCGCGCG 7958  
 QY 1298 --GGCTTGAATCGGGAAGATTATGAAACGTTGGGGGAGAGCTGTCTCAACTTTTCG 1355  
 DB 7957 CGGCGCCACACGCTGCTGATACATGCGCTGCTGTGCGAGGACTTGTGTAACCTGCGCG 7898  
 QY 1356 TCCGCTGCTGCGCTGAGCGCGCTGTGCGCTTGGCGCTTGGCGCGCATGCGGGCGG 1415  
 DB 7897 CGACCGCATCTCGGGCGCACTGCGCGCGCATGCTGCGCGCGCTTGGCGGGCGG 7838  
 QY 1416 GCGCGTGTCTGATTTGGCA-----GATGCGCGCTTCAAGCTTGGCGCGAGAG 1467  
 DB 7837 ACCCGTCCCGCTGATCGGCAACAGAGGGCGCACGACGCGCAGAGAGTGCAGCGCGCA 7778  
 QY 1468 CTGCGTTTGGCGCT-----CGTGGCAATTCGCTGCGCG 1502  
 DB 7777 CTTCGCGCATGCGCATCCGTCGCGCGTACCGCAAGCGCGCGCGATACGCGCTGCGCA 7718  
 QY 1503 CGAGTAAACCTGCGCATGCTGTCTCATCATGACACTTCGCGCGCGGATTTGCGAGGC 1562

DB 7717 CAAACTCGGTCTCCGCTGCTGACGTTCTGACACACCCCGGCGCTTACCCGCGCGGT 7658  
 QY 1563 GCGTAGAGAGCTTCGGATGCGAAGCTTCATTTGCGCGCACTTGTCCAACTATGACGC 1622  
 DB 7657 CGCGGAGAGAGGCGCGCGCGCGCTGCGCGCATGCGCGCAAGCTGCGCTATGCGCGGCT 7598  
 QY 1623 TCCCTCCCGCACTTTCGCTATTTATGTAAGGCGCTTGGCGGCGCGCGCTGCGCAT 1682  
 DB 7597 GCGCGTCCGCTGCTGACCGCTGTGTCACCGCGCGAGGCGCGCGCGCGCTGCGCT 7538  
 QY 1683 GCTGCCCGCGATCTGTCTACCGCGCGGAAACGCGTGTGCGCATTTGCGACACGA 1742  
 DB 7537 CGCCGCGCAACCGCTGCTGATGTTGAGAACGCGCATCTGCGTATACGCGCGCA 7478  
 QY 1743 GCGCGCTGCGCATCTTCTTCCGAGACACGCGCGCGGAAATATGAGCGACA 1802  
 DB 7477 GCGCTGCGCGCATCTGTGCGAGACCCCGCGCGCGCGCGCTGCGCGCGCGCT 7418  
 QY 1803 AGCGTGCAGGCGCGCACTTTAAGCCAGGCGCTTATGACGCGGATGTCGCGCGAAC 1862  
 DB 7417 GAGGTCACCGCGCGCACTCTGCGCTTGGCATGTCGAGCGCGCTTCCCGGAC 7358  
 QY 1863 CGAGCATTTTGTGAAGAAATTCGCGCAATGACCAAGCGCTTCCGA 1913  
 DB 7357 CGAAGCGCGCACCGGAGCGGACCGCGTGGCGCGCGCGCGCGCGCGCGCGCG 7307

RESULT 9  
 US-09-107-532A-170  
 ; Sequence 170, Application US/09107532A  
 ; Patent No. 6583275  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn A Ducelette-Stamm and David Bush  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 7310  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: GENOME THERAPEUTICS CORPORATION  
 ; STREET: 100 Beaver Street  
 ; CITY: Waltham  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02354  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD-ROM ISO9660  
 ; COMPUTER: PC  
 ; OPERATING SYSTEM: <Unknown>  
 ; SOFTWARE: ASCII  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/107,532A  
 ; FILING DATE: 30-Jun-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/085,598  
 ; FILING DATE: 14 May 1998  
 ; APPLICATION NUMBER: 60/051571  
 ; FILING DATE: July 2, 1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ariniello, Pamela Deneke  
 ; REGISTRATION NUMBER: 40,489  
 ; REFERENCE/DOCKET NUMBER: GTC-012  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (781)893-5007  
 ; TELEFAX: (781)893-8277  
 ; INFORMATION FOR SEQ ID NO: 170:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 870 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: circular  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:



TOPOLOGY: linear

Query Match 4.5%; Score 96.2; DB 4; Length 19702;  
Best Local Similarity 49.5%; Pred. No. 1.1e-17;  
Matches 248; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 686 ATGAATCGGTAAATTAATGAGAGAGGACCGGTGAGGAGCAATCCGATGAGCGTAAATTTTGT 745  
DB 11553 ATGAAGCGGTGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 11494  
QY 746 CCGATTTTCTTCTCGCGCGGTTCTTTGGGCAAGGTCGCGTGGGATGAGGAG 805  
DB 11493 TGGATTTCACTTTTATCATGAGGCTTCTATGAGTACGAGTGTAGGAGAAATCACTCGTT 11434  
QY 806 CGATTCACCGCGCCACAGAGCTGAAATCTCCCACTGCTGCTCCCTCGCTTCCGCTGAGT 865  
DB 11433 TGTTCAGTATCGAGCTGCGAATAATGCGAGTGTCTCTATTCACAGGCTCTGCTGAGAG 11374  
QY 866 CGCGCATGAGAGAGAGCAATCGAGCTTTTGTATGATGAGTGTATGATGATGATGATGATG 925  
DB 11373 CCGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 11314  
QY 926 AGCTTCACCGCGAGGAGCAATTCGCGTTCCTGTGTATTTGGCAATCCAGATGAGT 985  
DB 11313 AAGCGCATTCAGATGCTGCTGCTCTTTTACCTGACCATTTTGAAGATCCAGAGCTGAGT 11254  
QY 986 GCGCATGAGGCTCGTGGGTTTATCTGAGGATCTCACTTTTGGGAGAGCCGCGCGAGAG 1045  
DB 11253 GTGACAGAGCTTTCTTCTGATGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 11194  
QY 1046 TAGTTTCTCGGCTCTCGCTGAGTGAATTAACATGAGGAGGAGGAGGAGGAGGAGGAG 1105  
DB 11193 TTGTTTGTCTGGGCTCTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 11134  
QY 1106 TGCAGCAGCGGAGAGATTTGTAAGAACTGTGTGATGAGAGATTTGTGTCGCACTCC 1165  
DB 11133 TCAGAAAGGAGAGATTTCTATTAAGAACATGAGGCTTTGTGATGATCTATTTGCAAGAGAG 11074  
QY 1166 AATTGCGTGCAGCGGTGCGCA 1186  
DB 11073 ACTTACAGATACGATGCTA 11053

RESULT 12

US-09-557-884-1  
Sequence 1, Application US/09557884  
Patent No. 6506581

GENERAL INFORMATION:

APPLICANT: Fleischmann et al.  
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS: The Haemophilus influenzae Rd Genome, Fragments thereof, and Uses Thereof

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM: MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA: APPLICATION NUMBER: US/09/557,884

FILING DATE: 25-Apr-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB18693

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-309-8504

TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1  
Query Match 4.5%; Score 96.2; DB 4; Length 1830121;  
Best Local Similarity 49.5%; Pred. No. 1.2e-16;  
Matches 248; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 686 ATGAATCGGTAAATTAATGAGAGAGGACCGGTGAGGAGCAATCCGATGAGCGTAAATTTTGT 745  
DB 1336980 AAGATGCGCTAATTAATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1337039  
QY 746 CCGATTTTCTTCTCGCGCGGTTCTTTGGGCAAGGTCGCGTGGGATGAGGAG 805  
DB 1337040 CGAATTTTCTTCTTATGAGGCGGTTCAATGAGGCTTCTGTAGTGTGTCGCAAAATTTGTTAAG 1337099  
QY 806 CGATTCACCGCGCCACAGAGCTGAAATCTCCCACTGCTGCTCCCTCGCTTCCGCTGAGT 865  
DB 1337100 CGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1337159  
QY 866 CCGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 925  
DB 1337160 CTGATATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1337219  
QY 926 AGCTTCACCGCGAGGAGCAATTCGCGTTCCTGTGTATTTGGCAATCCAGATGAGT 985  
DB 1337220 CTCAGATGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1337279  
QY 986 GCGCATGAGGCTCGTGGGTTTATCTGAGGATCTCACTTTTGGGAGAGCCGCGCGAGAG 1045  
DB 1337280 GCGTATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1337339  
QY 1046 TAGTTTCTCGGCTCTCGCTGAGTGAATTAACATGAGGAGGAGGAGGAGGAGGAGGAG 1105  
DB 1337340 TTGTTTGTCTGGGCTCTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1337399  
QY 1106 TGCAGCAGCGGAGAGATTTGTAAGAACTGTGTGATGAGAGATTTGTGTCGCACTCC 1165  
DB 1337400 TCAGAGAGAGAGAGATTTCTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1337459  
QY 1166 AATTGCGTGCAGCGGTGCGCA 1186  
DB 1337460 AATGCGTCAAACTTTAGCA 1337480

RESULT 13

US-09-643-990A-1  
Sequence 1, Application US/09643990A  
Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann

Mark D. Adams

Owen White

Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.





**RESULT 15**

US-08-973-275-2

Sequence 2, Application US/08973275B

GENERAL INFORMATION

; GENERAL INFORMATION:

APPLICANT: MATSUNAGA, Tadashi

APPLICANT: NAMBA, KENTYO

TITLE OF INVENTION: FINE

TITLE OF INVENTION: BOUND THERETO, PROCESS FOR PRODUCING THE SAME, AND USE THEREOF

FILE REFERENCE: MATSINAGA

CURRENT APPLICATION NUMBER:

; CURRENT FILING DATE: 1998-02-09

EARLIER APPLICATION NUMBER: PCT/JP97/01043  
EARLIER FILING DATE: 1997-03-27

EARLIER APPLICATION NUMBER: JP 8-97536

EARLIER FILING DATE: 1996-03-28

EARLIER APPLICATION NUMBER: JP 8-146833  
EARLIER FILING DATE: 1995-05-16

NUMBER OF SEQ ID NOS: 6

```

; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 2
LENGTH: 954

```

```

:      LENGTH: 553
:      TYPE: DNA
:

```

**ORGANISM:** *Magnetospirillum*

**FEATURE:**

NAME/KEY: LOCATION:

OTHER INFORMATION:

OTHER INFORMATION: of the mps ge

OTHER INFO  
IS-08-973-275-3

7-017-010-0000

Query Match

Best Local Similarity 33.6%; Pred. No. 9.4e-16;  
 Matched 153; Combinations 74; W/ matched 205

```

Matches 152; conservative 54; mismatches 24; indels 0; gaps

```

QY 1458 C

Query Match	Score	DB 2;	Length
4.18;	87.4;		954;

Best Local Similarity 33.6%; Pred. NO. 9.4e-16;  
Matches 152; Conservative 54; Mismatches 247

[illegible]

QY	1458	GCCGACAGAGCTGGCTTTTGGCGCGTGGGACATTTTGGCTGGCGCGACGACTAACTGGC	151
Db	393	RCGCGAGGNTAAYMGNAAGCNAARMGYNTATGGARATGGCGNAYCAATTTCAAGTNC	452
QY	1518	GATGTGTCCATCATCCGACCTCCGGCGCGGAATTTGCGAGCGCGTGAAGAGCTCGG	1577
Db	453	NATATATACNTYNTINGAYACNGCNGGCGNTAYCCNGANGTNGAYGGCNGARGCWNMGNG	512
QY	1578	CATGCAAGCTCGATTGGCGGCGACCTGTTCCAGGTTATCCAGCGTCCCTCCCAACCGT	1633
Db	513	NCARGCNGARGCNAITHGCMGNMNSNATHGABRCNTGYTNAAYGTMNNGTNCNNTNGT	572
QY	1638	TTCCGTCATTATTGGTCAGGGCGTTGGCGGTGGCGCGCTGGCCATGCTGCCCGCGATCT	1697
Db	573	NMSGTNTATHTHGGNARGNGGNGMSGNGGNGCNAITHGNTYNTGNCNACNGNAATAC	632
QY	1698	GGTCTACGCGGCCGAAAACGCGTGGCTGTCCGATTGCCACGAGGGCGCCTCGGCAT	1757
Db	633	NGNTYNTATGYTNBARCAYGCNATHTAWSNGTNAITHMSNCNGARGNTGYCWNMSNAT	692
QY	1758	CCTCTTCCGCGACCAACACGCGCGGGAATCATTAAGAGCGAAGGCGGTGACGGCCCA	1811
Db	693	HYTNTGMMWSNCGNGBAAVGCNAAARGATGCGNCNGARCARHTNMENYTNACNGCNCA	752
QY	1818	CGCATTTTAAAGCCAAAGGCTTATCGACGGGATCGTCGCCGAAACCGAGCATTTGTGTA	1877
Db	753	RGAYTTCNCAABARTYTNMSNATHATGAWMSGTGTNCNGARCCMATGGGNGGNGCNCA	812
QY	1878	AGAAATTTCTGGGCACATCAGCAACGCCCTCTC	1910
Db	813	YMGNAAYCCNGAYYTNATGATGCARACNYYTMS	845

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 20:26:35 ; Search time 1275.22 Seconds  
(without alignments)  
5440.955 Million cell updates/sec

Title:	US-10-024-370-1
Perfect score:	2123
Sequence:	1 ctgcgagcggagtcgtrgat.....gaccttcattgcttgaaatc 2123

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing:	Minimum Match 0%
	Maximum Match 100%
	Listing first 45 summaries

```
Database : Published Applications NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*
```

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2123	100.0	2123	13	US-10-024-370-1	Sequence 1, Appl1
2	2111	99.4	3309400	10	US-09-738-626-1	Sequence 1, Appl1
3	1473	69.4	1473	13	US-09-738-626-929	Sequence 929, App
4	1473	69.4	1473	13	US-10-024-370-2	Sequence 2, Appl1
5	312.6	14.7	1488	10	US-09-712-363-35	Sequence 35, Appl1
6	131	6.2	354	10	US-09-738-626-3442	Sequence 3442, App
7	122.8	5.8	873	9	US-09-815-242-7800	Sequence 7800, App
8	121	5.7	867	9	US-09-815-242-6828	Sequence 6828, App
9	121	5.7	6021	10	US-09-070-9274-458	Sequence 458, App
10	119.2	5.6	870	9	US-09-815-242-7288	Sequence 7288, App
11	119.2	5.6	1053	10	US-09-895-913A-97	Sequence 97, Appl1
12	108	4.8	993	9	US-09-815-242-9854	Sequence 9854, App
13	102.8	4.1	915	9	US-09-815-242-6133	Sequence 6133, App
14	97.8	4.6	867	9	US-09-815-242-9520	Sequence 9520, App
15	96.8	4.6	654	10	US-09-974-300-5531	Sequence 5531, App
16	96.2	4.5	891	9	US-09-815-242-7115	Sequence 7115, App

17	96.2	4.5	1830121	14	US-10-322-960-1	Sequence 1, Appl1
18	93	4.4	867	9	US-09-815-242-9267	Sequence 9267, Ap
19	86.4	4.1	858	9	US-09-815-242-8520	Sequence 8520, Ap
20	85.2	4.0	756	9	US-09-815-242-4369	Sequence 4369, Ap
21	84.8	4.0	7972	8	US-08-781-9864-312	Sequence 312, App
22	81.8	3.9	991	9	US-09-815-242-7836	Sequence 7836, Ap
23	79	3.7	1335	14	US-10-156-761-5702	Sequence 5702, Ap
24	79	3.7	9025608	14	US-10-156-761-1	Sequence 1, Appl1
25	67.8	3.2	1596	14	US-10-156-761-3318	Sequence 3318, Ap
26	67.8	3.2	9025608	14	US-10-156-761-1	Sequence 1, Appl1
27	59.6	2.8	789	9	US-09-815-242-6829	Sequence 6829, Ap
28	59.6	2.8	798	9	US-09-815-242-3509	Sequence 3509, Ap
29	57.2	2.7	971	10	US-09-914-300-1085	Sequence 1085, Ap
30	56.4	2.7	7989	10	US-09-070-927A-269	Sequence 269, App
31	54.6	2.6	960	9	US-09-815-242-5951	Sequence 5951, Ap
32	52.6	2.5	960	9	US-09-815-242-9799	Sequence 9799, Ap
33	49.8	2.3	768	9	US-09-815-242-9268	Sequence 9268, Ap
34	49.8	2.3	768	9	US-09-815-242-9521	Sequence 9521, Ap
35	49.8	2.3	1128	14	US-10-156-761-3349	Sequence 3349, Ap
36	48.6	2.3	220	9	US-09-815-242-2856	Sequence 2856, Ap
37	48.6	2.3	1629	14	US-10-156-761-3330	Sequence 3330, Ap
38	46.2	2.2	984	9	US-09-767-479-11	Sequence 11, Appl1
39	46.2	2.2	618	13	US-10-003-446-14	Sequence 14, Appl1
40	46	2.2	975	9	US-09-841-132-461	Sequence 461, App
41	45.8	2.2	2225	9	US-09-815-242-4026	Sequence 4026, Ap
42	45	2.1	1125	14	US-10-156-761-2173	Sequence 2173, Ap
43	44.8	2.1	951	14	US-10-156-761-7101	Sequence 7101, Ap
44	44.4	2.1	72604	12	US-10-162-497-7	Sequence 7, Appl1
45	43.4	2.0	2101	12	US-10-289-757-151	Sequence 151, App

## ALIGNMENTS

```

RESULT 1
US-10-024-370-1
Sequence 1, Application US/10024370
Publication No. US20020142405A1
GENERAL INFORMATION:
APPLICANT: TILG, YVONNE
APPLICANT: ELKMANN, BERND
APPLICANT: EGGELING, LOTHAR
APPLICANT: SAHM, HERMANN
APPLICANT: MCKEL, BETTINA
TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY
TITLE OF INVENTION: FERMENTATION AND NUCLEOTIDE SEQUENCES CODING FOR THE
TITLE OF INVENTION: acdCA GENE
FILE REFERENCE: 21123-284139-MS
CURRENT APPLICATION NUMBER: US/10/024,370
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/362,899
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: DE 199 24 365.4
PRIOR FILING DATE: 1999-07-29
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.1

```

Query Match	Score	DB	Length
100.0%	2123	13	2123

QY 1 CTCGAGCGGGGAGTCGGTGATCGGCCACTCTCTTAAGCAATGCCGCGCTTTAAATTAAGCAA 600

Db 1 CTCGAGCGGGGAGTCGGTGATCGGCCACTCTTAAGCAATGCCGCGCTTTAAATTAAGCAA 600

OY	61	CTTAAATGTTTCTACCACTGTCGGCGACGACGACGAAAGTATGTTTTCATACACGTA	120
Db	61	CTTAAATGTTTCTACCACTGTCGGCGACGACGACGAAAGTATGTTTTCATACACGTA	120
OY	121	AACGTGTAAAGTGAAGTACCTAACCTCACAATTGCAATGCGAATTGGAACCTC	180
Db	121	AACGTGTAAAGTGAAGTACCTAACCTCACAATTGCAATGCGAATTGGAACCTC	180
OY	181	ACTCCCCCAATATCTTAACTTAACTTAAAGTAGTGTTTTACTGCAATTATTAAGT	240
Db	181	ACTCCCCCAATATCTTAACTTAACTTAAAGTAGTGTTTTACTGCAATTATTAAGT	240
OY	241	TCCCGATCTACCCCTCTTTACCCCGAAATACCCCTTTGCAAGAGTTGCAACACACA	300
Db	241	TCCCGATCTACCCCTCTTTACCCCGAAATACCCCTTTGCAAGAGTTGCAACACACA	300
OY	301	GTGCAATAGTAAACGGGCTTCAACAGTCCATTTCTCGGTTTAAAGCTATGTTCCGG	360
Db	301	GTGCAATAGTAAACGGGCTTCAACAGTCCATTTCTCGGTTTAAAGCTATGTTCCGG	360
OY	361	ACGCTAGGCAAAAGATGTTTGTGAGATGGAAGCAATATCCGCTATTTTAAACGAA	420
Db	361	ACGCTAGGCAAAAGATGTTTGTGAGATGGAAGCAATATCCGCTATTTTAAACGAA	420
OY	421	TGATAGGCTAAATTTGGGCTTAGATCTTCCGCTCTAAATAGTATGCAAGACATTCGA	480
Db	421	TGATAGGCTAAATTTGGGCTTAGATCTTCCGCTCTAAATAGTATGCAAGACATTCGA	480
OY	481	ATTATATTAAACAAGCCATTTTTCGCGCGTGGAGAAAGCGTTTCCGACTATGATGTGGGCG	540
Db	481	ATTATATTAAACAAGCCATTTTTCGCGCGTGGAGAAAGCGTTTCCGACTATGATGTGGGCG	540
OY	541	ATGGAACACACTTTCAGCATATGACGCTACATAGACTCGGTTTGTGAAACCTCGACAGCTTCAT	600
Db	541	ATGGAACACACTTTCAGCATATGACGCTACATAGACTCGGTTTGTGAAACCTCGACAGCTTCAT	600
OY	601	TCTTGGAAATGAACCTCCCAATATGACACCTCTAATCAAGGCTATGCAAGACTTTGAG	660
Db	601	TCTTGGAAATGAACCTCCCAATATGACACCTCTAATCAAGGCTATGCAAGACTTTGAG	660
OY	661	CGGGCTCGAAGCAAGGCCAAATGCGATGATGAGTAACTTACTGAGAAAGGCAACGCTGGAG	720
Db	661	CGGGCTCGAAGCAAGGCCAAATGCGATGATGAGTAACTTACTGAGAAAGGCAACGCTGGAG	720
OY	721	GGCATTTCCGGATGCCGTTATTTTGTCCGATTTTTCCTTCCTCGCGGTTCTTTGGGCAAG	780
Db	721	GGCATTTCCGGATGCCGTTATTTTGTCCGATTTTTCCTTCCTCGCGGTTCTTTGGGCAAG	780
OY	781	GTCCGGTGTGGTGGCATCATGAAAGCGATTACCGCGCCACAGAGCTGAATCTCCACCTG	840
Db	781	GTCCGGTGTGGTGGCATCATGAAAGCGATTACCGCGCCACAGAGCTGAATCTCCACCTG	840
OY	841	CTGCTCTCCCTGCTTCCGGTGTGCGCGCATGCAAGAAAGCAATTCGAGCTTTGTCAATG	900
Db	841	CTGCTCTCCCTGCTTCCGGTGTGCGCGCATGCAAGAAAGCAATTCGAGCTTTGTCAATG	900
OY	901	ATGATGTCCATAACCGCGGCTGTGACGCTCAACGCGAGCGCATTTTGCCTTCTGAGT	960
Db	901	ATGATGTCCATAACCGCGGCTGTGACGCTCAACGCGAGCGCATTTTGCCTTCTGAGT	960
OY	961	TATTTGGCGAATCCCAACGATGGGTGGCGCATAGGCTGTGGGGTTTCACTGGGATCTC	1020
Db	961	TATTTGGCGAATCCCAACGATGGGTGGCGCATAGGCTGTGGGGTTTCACTGGGATCTC	1020
OY	1021	ACTTTTGGCGAACCAGCGCGCGACGATAGGTTTCTGTGGTCTCTCGGCTGGTGAAGTTAAC	1080
Db	1021	ACTTTTGGCGAACCAGCGCGCGACGATAGGTTTCTGTGGTCTCTCGGCTGGTGAAGTTAAC	1080
OY	1081	ACTGGGCAATGCCCTTTCAGACGCTGTGCAGCAGCGGAGAAATTTGTGAAAATCTGTGTG	1140
Db	1081	ACTGGGCAATGCCCTTTCAGACGCTGTGCAGCAGCGGAGAAATTTGTGAAAATCTGTGTG	1140

QY	1141	ATTGATGGAAATTGGTGGCCACTCCATTTGGGTGACAGCGGTGGAAAAAACCCTCAAGTT	1200
Db	1141	ATTATATGAATTTGGTGGCCACTCCAAATGGGTGACACGGGTGGAAAAAACCCTCAAGTT	1200
QY	1201	ATTGACCGGTAGAGGCAACGATCGATTTTCTTCAACAACTCCTGGCGTGGCACTTCG	1260
Db	1201	ATTGACCGGTAGAGGCAACGATCGATTTTCTTCCAAACAACTCCTGGCGTGGCACTTCG	1260
QY	1261	GTGATGAGAGCGAATTGGCGCTTCTCTGTGACCCGCAAGGCGCTGGATTCGGGAGATTATG	1320
Db	1261	GTGATGAGAGCGAATTGGCGCTTCTCTGTGACCCGCAAGGCGCTGGATTCGGGAGATTATG	1320
QY	1321	GAAACGTTGGGGGGGAGACGTCGTCAACGTTCTGTGTGCGGTGCTGTGGCGAATTGAGGCG	1380
Db	1321	GAAACGTTGGGGGGGAGACGTCGTCAACGTTCTGTGTGCGGTGCTGTGGCGAATTGAGGCG	1380
QY	1381	GCTGTGCGCTGTGGCTTCGCGCGCATCGAGGAGCGGCCCGCTGTGTCTGATTGGCGAGAT	1440
Db	1381	GCTGTGCGCGGTGTGCTTCGCGCGCATCGAGGAGCGGCCCGCTGTGTCTGATTGGCGAGAT	1440
QY	1441	CGCCGCTTCAAGCTTGGGCGCGAGAGCTGGCTTTGGCGGTCTGTGCATTTGCGTGGCG	1500
Db	1441	CGCCGCTTCAAGCTTGGGCGCGAGAGCTGGCTTTGGCGGTCTGTGCATTTGCGTGGCG	1500
QY	1501	CGCGAGCTAAACCTGCGCGATCGATGTCATATCGACACTTCGCGCGCGCAATTGTCAAGCTTATGCAC	1560
Db	1501	CGCGAGCTAAACCTGCGCGATCGATGTCATATCGACACTTCGCGCGCGCAATTGTCCGAG	1560
QY	1561	GCGGCTGAGAGCTCGGATCGCAAGCTCGATTGGCGCACTTGTCAAGCTTATGCAC	1620
Db	1561	GCGGCTGAGAGCTCGGATCGCAAGCTCGATTGGCGCACTTGTCAAGCTTATGCAC	1620
QY	1621	GCTCCCTTCCCAACCGTTTGGGTATTATTTGTAGAGGCGTTGGCGGTGGCGCGCTTGACC	1680
Db	1621	GCTCCCTTCCCAACCGTTTGGGTATTATTTGTAGAGGCGTTGGCGGTGGCGCGCTTGACC	1680
QY	1681	ATGCTGCCCGCGCATCTGTCTTAGCGCGGCGAACCCTGTGCTGTCCGATTCACACA	1740
Db	1681	ATGCTGCCCGCGCATCTGTCTTAGCGCGGCGAACCCTGTGCTGTCCGATTCACACA	1740
QY	1741	GAGGCGGCTTGGGCATCTCTTCCGCGACAACCAACGAGCGCGGAAATCATAGACGA	1800
Db	1741	GAGGCGGCTTGGGCATCTCTTCCGCGACAACCAACGAGCGCGGAAATCATAGACGA	1800
QY	1801	CAAGCGGTGACGGGCAACGCACTTTTAAGCCAAAGGCTTATCCAGCGGATTCGTCCGAA	1860
Db	1801	CAAGCGGTGACGGGCAACGCACTTTTAAGCCAAAGGCTTATCCAGCGGATTCGTCCGAA	1860
QY	1861	ACCGGACCTTGTTTGAAAGAAATTCCTGGGCAATCAGCAACGCGCTCTCGAATTGGAT	1920
Db	1861	ACCGGACCTTGTTTGAAAGAAATTCCTGGGCAATCAGCAACGCGCTCTCGAATTGGAT	1920
QY	1921	AACCAATCCGAGAGAGGCGGAGCGGACAAGTCGCTTACACGATTTGAGCGTTTAGCGCAG	1980
Db	1921	AACCAATCCGAGAGAGGCGGAGCGGACAAGTCGCTTACACGATTTGAGCGTTTAGCGCAG	1980
QY	1981	TAAAGAAATTTATCCGCTGATCAATGATGATGATGATAACACAGGTTAGGCGCAAGCATTGG	2040
Db	1981	TAAAGAAATTTATCCGCTGATCAATGATGATGATGATAACACAGGTTAGGCGCAAGCATTGG	2040
QY	2041	GTGGCGCGAAACCTTCAGGCGCGTAAAGAGCCTTCGCGGAAATGGTCAAGTCAACGATCC	2100
Db	2041	GTGGCGCGAAACCTTCAGGCGCGTAAAGAGCCTTCGCGGAAATGGTCAAGTCAACGATCC	2100
QY	2101	GCCGACCTTACGCTTGAATTC	2123
Db	2101	GCCGACCTTACGCTTGAATTC	2123

RESULT 2  
US-09-738-626-1/c  
; Sequence 1, Application US/09738626  
; Publication No. US20020197605A1

GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 98/37484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 99/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: Patentin ver. 3.0  
SEQ ID NO 1  
LENGTH: 3309400  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

Query Match 99.4%; Score 2111; DB 10; Length 3309400;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2122; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CTCGAGCGGAGTCCGCTGATCGGCCACTCTCTAGCAATGCGGCTTTAAATAAGCA 60  
DB 881620 CTCGAGCGGAGTCCGCTGATCGGCCACTCTCTAGCAATGCGGCTTTAAATAAGCA 881561

QY 61 CTTATATGTTCTACCAACATCTGCGCGAGCAACAGAGTAATGTTGTCAGACAGCTA 120  
DB 881560 CTTATATGTTCTACCAACATCTGCGCGAGCAACAGAGTAATGTTGTCAGACAGCTA 881501

QY 121 AACGTGTAATGTAAGTAACTTAACCTCACTGCAATTCGATGCGATTTGGAACATC 180  
DB 881500 AACGTGTAATGTAAGTAACTTAACCTCACTGCAATTCGATGCGATTTGGAACATC 881441

QY 181 ACTCCCCCAATATCTTAATCTTAATAAGTAAGTGTCTTAACTGCAATTAATAAGT 240  
DB 881440 ACTCCCCCAATATCTTAATCTTAATAAGTAAGTGTCTTAACTGCAATTAATAAGT 881381

QY 241 TCCCGATCTACCCCTCTTACCCGAAATACCCCTTTGCAAGATTCGAAACACACAA 300  
DB 881380 TCCCGATCTACCCCTCTTACCCGAAATACCCCTTTGCAAGATTCGAAACACACAA 881321

QY 301 GTGCAATAGTAACGGGCTTCAACGTCACCAATCTGTCGGGTTTAGGCTATGTTGCGG 360  
DB 881320 GTGCAATAGTAACGGGCTTCAACGTCACCAATCTGTCGGGTTTAGGCTATGTTGCGG 881261

QY 361 ACCTCTAGGCAAAAGTAAGTGTGAGTAAGAAAGCATTAATCCGCTATTTTAAAGCAA 420  
DB 881260 ACCTCTAGGCAAAAGTAAGTGTGAGTAAGAAAGCATTAATCCGCTATTTTAAAGCAA 881202

QY 421 TCGATAGCTTAATAGGCTTGAATCTTCCGCTCTTAATAGTATGCAAGACATTCGA 480  
DB 881201 TCGATAGCTTAATAGGCTTGAATCTTCCGCTCTTAATAGTATGCAAGACATTCGA 881142

QY 481 ATTAAATTAACAAGCATTTTTCGCGCGTGAAGAGCGTTTTCGCACTATGTTGCGGCG 540  
DB 881141 ATTAAATTAACAAGCATTTTTCGCGCGTGAAGAGCGTTTTCGCACTATGTTGCGGCG 881082

QY 541 ATGAACACACTTCAGCATTTGACGTCATAGACTCGGTTTGGACCTGACAGCTTCATT 600  
DB 881081 ATGAACACACTTCAGCATTTGACGTCATAGACTCGGTTTGGACCTGACAGCTTCATT 881022

post deleted

QY 601 TCTTGAAATGAAGTCCCAATATGACAACTCAATCAAGGCTATGACAGACCTTGAG 660  
DB 881021 TCTTGAAATGAAGTCCCAATATGACAACTCAATCAAGGCTATGACAGACCTTGAG 880962

QY 661 CGGGCTCGAAGCAAGGCCCAATGCGATGCAATGCTTAATTAATGAGAAAGCAACGTTGAG 720  
DB 880961 CGGGCTCGAAGCAAGGCCCAATGCGATGCAATGCTTAATTAATGAGAAAGCAACGTTGAG 880902

QY 721 GGCATTTCCGGTACCGGTTATTTTGTCCGATTTTCTCTCGGCGGTTCTTTGGACAG 780  
DB 880901 GGCATTTCCGGTACCGGTTATTTTGTCCGATTTTCTCTCGGCGGTTCTTTGGACAG 880842

QY 781 GTCGCGTGGTGCAGATCATGAAAGGAGATTCACCGGCGCACAGAGCTGAATCCCACTG 840  
DB 880841 GTCGCGTGGTGCAGATCATGAAAGGAGATTCACCGGCGCACAGAGCTGAATCCCACTG 880782

QY 841 CTGGTCTCCCTGCTTCGCGTGTGCGCGCATGACAGAAAGACATGACCTTTTGTGATG 900  
DB 880781 CTGGTCTCCCTGCTTCGCGTGTGCGCGCATGACAGAAAGACATGACCTTTTGTGATG 880722

QY 901 ATGGTGTCTATACCGCGGCTGTGACAGGTCAACCGGAGGCGCATTTGCTCTGCTG 960  
DB 880721 ATGGTGTCTATACCGCGGCTGTGACAGGTCAACCGGAGGCGCATTTGCTCTGCTG 880662

QY 961 TATTTGCGCAATCCCAAGATGAGGAGGCGCATGAGGCTGAGGCTTCAATGAGGATCTC 1020  
DB 880661 TATTTGCGCAATCCCAAGATGAGGAGGCGCATGAGGCTGAGGCTTCAATGAGGATCTC 880602

QY 1021 ACTTTTGGGAAACCCGCGCGCGAGATAGGTTTCTGAGGCTCTGCGGTGTGAGTTAAC 1080  
DB 880601 ACTTTTGGGAAACCCGCGCGCGAGATAGGTTTCTGAGGCTCTGCGGTGTGAGTTAAC 880542

QY 1081 ACTGGGCAATGCGCTTCCAGACGCTGTGACAGCGGCGGAGAAATTTGGTGAACCTGCTG 1140  
DB 880541 ACTGGGCAATGCGCTTCCAGACGCTGTGACAGCGGCGGAGAAATTTGGTGAACCTGCTG 880482

QY 1141 ATTGAATGAAATGATGCGCACTCAATTTGAGTGAAGGAGGAGGAGGAGGAGGAGG 1200  
DB 880481 ATTGAATGAAATGATGCGCACTCAATTTGAGTGAAGGAGGAGGAGGAGGAGGAGG 880422

QY 1201 ATTCAAGCGGTGAGAGCAAGCATGCTTTTCTCAACAACCTCTGAGGCTGACATTCG 1260  
DB 880421 ATTCAAGCGGTGAGAGCAAGCATGCTTTTCTCAACAACCTCTGAGGCTGACATTCG 880362

QY 1261 GTGATGAGGCGATTCGCGCTTCTGTGACCCGAGAGGCTGTGAATGAGGAGATTAATG 1320  
DB 880361 GTGATGAGGCGATTCGCGCTTCTGTGACCCGAGAGGCTGTGAATGAGGAGATTAATG 880302

QY 1321 GAAAGCTTGGGGGAGAGAGTGTCAAGCTTTCTGAGGCGGCTGAGGCGCATTAAGCCG 1380  
DB 880301 GAAAGCTTGGGGGAGAGAGTGTCAAGCTTTCTGAGGCGGCTGAGGCGCATTAAGCCG 880242

QY 1381 GCTGTGCGCGTACCTGCGCGCATCGGAGGCGCGGCGGCTGTGATGAGGAGAT 1440  
DB 880241 GCTGTGCGCGTACCTGCGCGCATCGGAGGCGCGGCGGCTGTGATGAGGAGAT 880182

QY 1441 GCGCGCTTACGCTTGGGCGCGAGAGCTGCTTTTGGCGGTGTGAGCATTTTGGCTGCG 1500  
DB 880181 GCGCGCTTACGCTTGGGCGCGAGAGCTGCTTTTGGCGGTGTGAGCATTTTGGCTGCG 880122

QY 1501 CGCGAGCTAAACCTGCGCATGTGTCATTCGACACTTCGCGGCGGAGATTTGCGAG 1560  
DB 880121 CGCGAGCTAAACCTGCGCATGTGTCATTCGACACTTCGCGGCGGAGATTTGCGAG 880062

QY 1561 GCGGCTGAGAGCTTCGAGTGCAGATTCGATTCGAGGCGCACCTTGTCCAGCTTAATGAC 1620  
DB 880061 GCGGCTGAGAGCTTCGAGTGCAGATTCGATTCGAGGCGCACCTTGTCCAGCTTAATGAC 880002

QY 1621 GCTCCCTTCCCAACCGTTTGGGCTATTAATGCTACGAGGCGTTGGCGGTGGCGGCTGCG 1680  
DB 880001 GCTCCCTTCCCAACCGTTTGGGCTATTAATGCTACGAGGCGTTGGCGGTGGCGGCTGCG 879942

QY 1681 ATGCTGCGCGCGATCTGCTTACGCGGCGGAGAAACGCGTGGCTGTCCGATTCGACCA 1740

Db	879941	ATGCTGCGCCGCGAATCTGCTCTACAGGGCGCAAAACGCGTGCCTCTCCGATTGCAACA	879882
QY	1741	GAGGGCGCCTCGGGCAATCTCTTCGCGCAACCAACGCGCGCGGAATCATAGAGCGA	1800
Db	879881	GAGGGCGCCTCGGGCAATCTCTTCGCGCAACCAACGCGCGCGGAATCATAGAGCGA	879822
QY	1801	CAAGCGTGCAGGCGCAGCACTTTTAAAGCAAGGCTTATCGACGGGATCGTGCAGAA	1860
Db	879821	CAAGCGTGCAGGCGCAGCACTTTTAAAGCAAGGCTTATCGACGGGATCGTGCAGAA	879762
QY	1861	ACCGAGCACTTTGTTGAAGAAATTCCTGGGCAATGAGCAAGCCTCTCCGATTTGAT	1920
Db	879761	ACCGAGCACTTTGTTGAAGAAATTCCTGGGCAATGAGCAAGCCTCTCCGATTTGAT	879702
QY	1921	AACATTCGGAAGAGGGCGGGAAGCGCAATCGCTTCAACAGATTTAGAGCTTAAAGGCAG	1980
Db	879701	AACATTCGGAAGAGGGCGGGAAGCGCAATCGCTTCAACAGATTTAGAGCTTAAAGGCAG	879642
QY	1981	TAAAGAAATTTATGCGCTGATCAATCGATGATGAACACCAAGGGTACGGCCAGACAGTGG	2040
Db	879641	TAAAGAAATTTATGCGCTGATCAATCGATGATGAACACCAAGGGTACGGCCAGACAGTGG	879582
QY	2041	GTGGCGGAAACCTTCAGGCGGTAAAGACGCTTCGCGGAATGTCAGCTGACGAGCTCC	2100
Db	879581	GTGGCGGAAACCTTCAGGCGGTAAAGACGCTTCGCGGAATGTCAGCTGACGAGCTCC	879522
QY	2101	GCCGACCTTCATGCTCGTAATTC	2123
Db	879521	GCCGACCTTCATGCTCGTAATTC	879499

RESULT 3  
US-09-73

```

: Sequence 929, Application US/09738626
: Publication No. US20020197605A1
:
: GENERAL INFORMATION:
: APPLICANT: NAKAGAWA, SATOSHI
: APPLICANT: MIZOGUCHI, HIROSHI
: APPLICANT: ANDO, SEIKO
: APPLICANT: HAYASHI, MIKIRO
: APPLICANT: OCHIAI, KEIKO
: APPLICANT: YOKOI, HARUHIKO
: APPLICANT: TATEISHI, NAOKO
: APPLICANT: SENOH, AKIHIRO
: APPLICANT: IKEDA, MASATO
: APPLICANT: OZAKI, AKIO
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-125
: CURRENT APPLICATION NUMBER: US/09/738,626
: CURRENT FILING DATE: 2000-12-18
: PRIOR APPLICATION NUMBER: JP 99/377484
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: JP 00/159162
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: JP 00/280988
: PRIOR FILING DATE: 2000-08-03
: NUMBER OF SEQ ID NOS: 7059
: SOFTWARE: PatentIn ver. 3.0
: SEQ ID NO 929
: LENGTH: 1473
: TYPE: DNA
: ORGANISM: Corynebacterium glutamicum
: US-09-738-626-929

```

	Query Match	Similarity	69.4%	Score 1473	DB 10	Length 1473
Best Local	Similarity	100.0%	Pred. No. 0			
Matches 1473	Conservative	0	Mismatches 0	Indels 0	Gaps 0	
Qy	508	GTGGAGAGCGTTTTCGACCTATGCTGTGGGGGATGAAACACACTTCAGCATTTGACGCTC	567			
Db	1	GTGGAGAGCGTTTTCGACCTATGCTGTGGGGGATGAAACACACTTCAGCATTTGACGCTC	60			

QY	568	ATPAGCTCGGTTTTGGAACCTCGACAGCTTCAATTTCTTGGAATGAATCTCCCAATATGAC	627
Db	61	ATAGACTCGGTTTTGGAACCTCGACAGCTTCAATTTCTTGGAATGAATCTCCCAATATGAC	120
QY	628	AACCTCAATCAAGGCTATGACAGACCTTTGGAGGGGCTCGAACAAGGCCAAATGGAT	687
Db	121	AACCTCAATCAAGGCTATGACAGACCTTTGGAGGGGCTCGAACAAGGCCAAATGGAT	180
QY	688	GAATCGGTAATTACTGAGAGAGGACCGTGGAGGGCAATCCGATAGCGTTATTTTGTCC	747
Db	181	GAATCGGTAATTACTGAGAGAGGACCGTGGAGGGCAATCCGATAGCGTTATTTTGTCC	240
QY	748	GATTTTTCCTTCTCTCGGCGGTTCTTTTGGGACCGGTCCGCTGGTGGCGCATATGAAGCG	807
Db	241	GATTTTTCCTTCTCTCGGCGGTTCTTTTGGGACCGGTCCGCTGGTGGCGCATATGAAGCG	300
QY	808	ATTACCGCGCGCACAGAGCTGAACTCCCACTCGTGGATCTCCCGTCCCGGATGAGTGGCG	867
Db	301	ATTACCGCGCGCACAGAGCTGAACTCCCACTCGTGGATCTCCCGTCCCGGATGAGTGGCG	360
QY	868	CGCATGCGAGAGAACAAATCGAGCTTTTGTATGATGATGTCATTAACCGCGCGCTGGCAG	927
Db	361	CGCATGCGAGAGAACAAATCGAGCTTTTGTATGATGATGTCATTAACCGCGCGCTGGCAG	420
QY	928	CGTCAACCGCGAGCGCATTTTCCGTTCTCTGTGTATTTTGGCAATCCACGATGGTGGC	987
Db	421	CGTCAACCGCGAGCGCATTTTCCGTTCTCTGTGTATTTTGGCAATCCACGATGGTGGC	480
QY	988	GCCATGGCGCTGATGGGGTTCAATCGGGGCAATCACTTTGGGAAACCGGAGGGCAGATA	1047
Db	481	GCCATGGCGCTGATGGGGTTCAATCGGGGCAATCACTTTGGGAAACCGGAGGGCAGATA	540
QY	1048	GATTTCCTGGGCTCCGCGTGGTGGAGTTAACCACTGGGCAATGCGCTTCCAGACGATGTG	1107
Db	541	GATTTCCTGGGCTCCGCGTGGTGGAGTTAACCACTGGGCAATGCGCTTCCAGACGATGTG	600
QY	1108	CAGCAGCGCGAGAAATTTGGTGAATACTGGTGTATTTGATGGAATTTGTGGCCACTCCA	1167
Db	601	CAGCAGCGCGAGAAATTTGGTGAATACTGGTGTATTTGATGGAATTTGTGGCCACTCCA	660
QY	1168	TTGGGTGAGCGGTGGGCAAAACCTCAAGTTTATTCAGCCGCGTAGAGGCAAGGATCGT	1227
Db	661	TTGGGTGAGCGGTGGGCAAAACCTCAAGTTTATTCAGCCGCGTAGAGGCAAGGATCGT	720
QY	1228	TTTTCTCCAAACAATCTCGTGGCGTGGCACTTCCGCTGTATGAGAGCGAATTTGCGCTTCTCGT	1287
Db	721	TTTTCTCCAAACAATCTCGTGGCGTGGCACTTCCGCTGTATGAGAGCGAATTTGCGCGTCTTCTCGT	780
QY	1288	GACCCGCGAGAGCGCTGGAATGCGGAGAAATTAATGGAACGTTGGGGCGACAGCTGTCAAG	1347
Db	781	GACCCGCGAGAGCGCTGGAATGCGGAGAAATTAATGGAACGTTGGGGCGACAGCTGTCAAG	840
QY	1348	CTTTCTGGGTGGCGGTGGCGGGAATTAAGCCCGGCTGGCGCGTGTGGCCCTGGCGGATC	1407
Db	841	CTTTCTGGGTGGCGGTGGCGGGAATTAAGCCCGGCTGGCGCGTGTGGCCCTGGCGGATC	900
QY	1408	GGGGGCGGCGCCGTGTGTCTGATTTGGGACAGGATGCGCGCTTCAAGCTTTGGGCGCAGAG	1467
Db	901	GGGGGCGGCGCCGTGTGTCTGATTTGGGACAGGATGCGCGCTTCAAGCTTTGGGCGCAGAG	960
QY	1468	CTGGCGTTTGGCGGTCTGGGCAATTTGCGTGGCGCGCAGACTTAACTTGGCGATGTGTCC	1527
Db	961	CTGGCGTTTGGCGGTCTGGGCAATTTGCGTGGCGCGCAGACTTAACTTGGCGATGTGTCC	1020
QY	1528	ATCATTCACACACTTCCGCGCGCTGAATTTGTTCGAGAGCGGCTAGAGAGCTTCGGATTCGACAGC	1587
Db	1021	ATCATTCACACACTTCCGCGCGCTGAATTTGTTCGAGAGCGGCTAGAGAGCTTCGGATTCGACAGC	1080
QY	1588	TCGATTTGGGCGCACTTGTCCCAACTTATGACGCTCCCTCCCGACCGTTTGGGTCAATT	1647
Db	1081	TCGATTTGGGCGCACTTGTCCCAACTTATGACGCTCCCTCCCGACCGTTTGGGTCAATT	1140
QY	1648	ATTGATCAGGGCGTTTGGCGGTGGCGACCTGGCCATGTCCCGCGAATGTGTTACGCG	1707



Db	1141	ATTGGTACGGGCGTGGCGGTGGCGCGCTGGCCATGCTGCCCGCATCTGGTCTACCGG	1200
Qy	1708	GCCGAAAAACGCGTGGCTGTCCGCATTGSCAACCAAGAGGGCGCCTTGCCCATCTCTTCCGC	1767
Db	1201	GCCCAAAAACGCGTGGCTGTCCGCATTGSCAACAGAGGGCGCCTTGCCCATCTCTTCCGC	1260
Qy	1768	GACACCAACCAACGCGCGCGAANAATATNAGAGCAAGCGCGTGCAGCGGCACAGCACTTTTA	1827
Db	1261	GACACCAACCAACGCGCGCGAANAATATNAGAGCAAGCGCGTGCAGCGGCACAGCACTTTTA	1320
Qy	1828	AGCCAAAGGCGCTTATCGACGGAGTGTGCGCCGAAACCGAGCACTTTGTTGAAGAATTCTC	1887
Db	1321	AGCCAAAGGCGCTTATCGACGGAGTGTGCGCCGAAACCGAGCACTTTGTTGAAGAATTCTC	1380
Qy	1888	GGCACATCAGCAACGCCCTCTCCGAATTGGATTAACAATCCGGAGAGGGCGGACCGGAC	1947
Db	1381	GGCACATCAGCAACGCCCTCTCCGAATTGGATTAACAATCCGGAGAGGGCGGACCGGAC	1440
Qy	1948	AGTCGCTTCAACAGCATTTGAGCGTTTAGCGAG	1980
Db	1441	AGTCGCTTCAACAGCATTTGAGCGTTTAGCGAG	1473

Db 1261 GACACCAACACGCGCGGAATCATPAGAGCAGAGGCGTGCAGCGCAGCACTTTTA 1320  
QY 1828 AGCCAGGCGTTATGAGGGGATCGTCGCCGAACCGAGCACTTTGTGAAGAAATTTCTC 1887  
Db 1321 AGCCAGGCGTTATGAGGGGATCGTCGCCGAACCGAGCACTTTGTGAAGAAATTTCTC 1380  
QY 1888 GGCACATCAGCAAGCCCTCTCCGAATTGATACATCCGAGAGAGGCGGAGCGCAGC 1947  
Db 1381 GGCACATCAGCAAGCCCTCTCCGAATTGATACATCCGAGAGAGGCGGAGCGCAGC 1440  
QY 1948 AGTCGCTTACACGATTTGAGCGTTAGCGCAG 1980  
Db 1441 AGTCGCTTACACGATTTGAGCGTTAGCGCAG 1473  
  
RESULT 5  
US-09-712-363-35  
; Sequence 35, Application US/09712363  
; Patent No. US20020164588A1  
; GENERAL INFORMATION:  
; APPLICANT: Eisenberg, David  
; APPLICANT: Rotstein, Sergio H.  
; APPLICANT: Marcotte, Edward M.  
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
; FILE REFERENCE: 07419-032001  
; CURRENT APPLICATION NUMBER: US/09/712,363  
; PRIOR FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: PCT/US00/02246  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/179,531  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 60/117,844  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: 60/118,206,  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: 60/126,593  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/134,093  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/134,092  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/165,124  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/165,086  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 35  
; LENGTH: 1488  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; US-09-712-363-35  
  
Query Match 14.7%; Score 312.6; DB 10; Length 1488;  
Best Local Similarity 54.0%; Pred. No. 2,1e-92;  
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

QY 812 ACCGCGCCACAGAGCTGAAATCTCCCACTGCTGCTCTCCCTGCTTCCGGTGTGCGCGCA 871  
Db 272 AGCGGGGACCGCGGACCGGCTGCGGCTACTGCGGTACCAAGCTCGAGGGACCGCGCA 331  
QY 872 TGCAAGAAAGCAATCGAGCTTTTGTGATGATGAGTGTCCATTAACCGCGGCTGTGAGCGTC 931  
Db 332 TGCAAGAAAGCAATCGAGCTTTTGTGATGATGAGTGTCCATTAACCGCGGCTGTGAGCGTC 391  
QY 932 ACCGCGAGGCGCAATTTCCGCTTCTGCTGTATTTTGGCAATCCCAATGAGTGTGCGCA 991  
Db 392 ACAACAGGCGCGCTCCCTACCTGCTATTTTGGCCATCCGACACCGAGTGTGAGTTT 451  
QY 992 TGCGCTGTGGGGTTCATCTGGGCAATCTCACTTTTGGGAAACCGGCGGCGAGTAGGTT 1051  
Db 452 TGCGCTGTGGGGTTCATCTGGGCAATCTCACTTTTGGGAAACCGGCGGCGAGTAGGTT 511  
QY 1052 TCCTGGGTCTCGCGTGTGAGTTAACTGAGGCACTGCGCTTCCAGAGGCTGTGAGC 1111  
Db 512 TTCGAGACCAAGGCTCTATGAGTGTCTATGAGCAACCTTCCATCCGGGCTCCAA 571  
QY 1112 AGCGGAGAAATTTGTGAAAACTGCTGTATGATGAAATTTGTGCGCACTCCATTTGC 1171  
Db 572 CCGCGGAAATCTACGGCGGCAATGAGATCATCGACGCGCTGTGCACTGAGCCGCTAC 631  
QY 1172 GTGAGCGGTGGCAAAACCTCAAGTTATTCAGCGCGTAGAGGCAACGATGCTTTT 1231  
Db 632 GACGATGCTGATGCTGCTGAGGCTGATGAGCGCTCCGAAACCGTTTCGCGAC 691  
QY 1232 CTCGAACAACCTCTGCGGTGCA---CTTCCGATGATGAGAGCGATTCGCGTCTGTG 1288  
Db 692 CGCAGAGCGCGCGCGCTGACCGATGCGCACTGAGGACTGGGTGTGAGCATCGCGCC 751  
QY 1289 ACCGCGAGAGCTGAGATGAGGAGATTTATGAAACGTTGGGAGCGAGAGTGTCAAGC 1348  
Db 752 GCGCGAGCGCGCGCGGTGAGGAGCTACTGCGACACGCGCCACCGAGCGAGGTGTTGT 811  
QY 1349 TTTTGGGCGCGCTGCTGCGCATTTAGCCCGGTGCGCGTTCGCGCGCGCATTCG 1408  
Db 812 TGTGAGAAACGATCAAGAGGAGCG---GGACACGCTGCTGCGCGCTGCGCGCTTTG 868  
QY 1409 GGGCGCGCGCGGTGCTGATTTGGGCGAG-----GATCGCGCTTCAAGC 1453  
Db 869 GCGGCGCAACCGAGTGTCTGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 928  
QY 1454 TTGGCGCGCAGAGCTCGTTTTCGCGTGTGCAATTTGCTGCGCGCGAGCTTAAC 1513  
Db 929 TCGGCGCGCGCTGCTTACGCGAAGCCGAGCGCGGAGTGTGCGCGCGAGCTGTGCC 988  
QY 1514 TGCGGATGCTGCTCATGATGAGCACTTCGCGCGCGCAATTTGTGCGAGGCGGTGAGAGC 1573  
Db 989 TGCGGCTGTGCTGTATGAGCGCGCGAGCTCGCGGTGTGCGCGCGAGCGAGCAGG 1048  
QY 1574 TCGGCAATCGAAGCTCGATTTGCGCGCACTTGTCCAACTTATGAGCGTCCCTCCCA 1633  
Db 1049 GCGGCGTGTGCGCGCAATGCGCATTTGCGCGAGGCTGTGAGCTGATGATGATGATGATG 1108  
QY 1634 CCGTTTGTGCTATTTGTGAGGCGTTGCGCGTGTGCGCGCTGTGCGCATGTGCGCGCG 1693  
Db 1109 CCGTGTGATCTGTGCGCGCAGGCGCGCGCGCGCGCGCGCGCGCGCATGTGCGCGCG 1168  
QY 1694 ATCTGTCTACGCGCGCGAAGAACCGGTGCTGTGCGCATTTGCGCAACGAGAGGCGCG 1753  
Db 1169 ACCGCGTGTGCGCGCACTCCAGCGGTGCGCGCTTGTCTTCCGAAAGAGACGCG 1228  
QY 1754 CCATCTCTTTCGCGAGACCAACCAACCGCGCGGAAATCATAGAGCAAGCGCTGTGAGG 1813  
Db 1229 CGATGCTGTTCGAGAGACGCTCATGCGCGCACTGTGCGCGCGCGCGCATTCGGT 1288  
QY 1814 CGACGCACTTTTAAGCAAGGCTTATGACGAGATGTGCGCGAAGCGAGCATTTTG 1873  
Db 1289 CCGCGGACCTACTGAAATCGGGGATGTGACACCATGCTGTGCGGAGTACCCGACGCG 1348

QY 1874 TTGAAGA 1880  
Db 1349 CAGACGA 1355

## RESULT 6

US-09-738-626-3442/c  
; Sequence 3442, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OKADA, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 3442  
; LENGTH: 354  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3442

Query Match 6.2%; Score 131; DB 10; Length 354;  
Best Local Similarity 100.0%; Pred. No. 2.1e-32;

Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1993 TGGCTGATCAATGATGATGAACACGAGGTACGCGCAGACAGTGGGTGCGCGAACC 2052  
Db 354 TGGCTGATCAATGATGATGAACACGAGGTACGCGCAGACAGTGGGTGCGCGAACC 295  
QY 2053 CTCAGGCGCGTAGAGGCTCTGCGCGAATGCTGAGCTGACGACGTCGCGACCTTCAT 2112  
Db 294 CTCAGGCGCGTAGAGGCTCTGCGCGAATGCTGAGCTGACGACGTCGCGACCTTCAT 235  
QY 2113 GCGTGGAAATTC 2123  
Db 234 GCGTGGAAATTC 224

## RESULT 7

US-09-815-242-7800  
; Sequence 7800, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zykend, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7800  
; LENGTH: 873  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(873)  
US-09-815-242-7800

Query Match 5.8%; Score 122.8; DB 9; Length 873;  
Best Local Similarity 52.8%; Pred. No. 2.1e-29;

Matches 265; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

QY 686 ATGAATCGGTAATTAATGAGAGGACACCGGAGGATTCGGTATGATTTGT 745  
Db 326 AGGAGCGCTGATGCGATGAGCGGCAAGCTGAGGGATGCGGTGCTGCGCT 385  
QY 746 CCGATTTTCTTCTGCGCGGCTTTTGGGACGCTGCGTGGTGCATGAAAG 805  
Db 386 TCGAGTTCTCTTCAATGAGCGGCTTCAATGAGGCGCATGCGCGGCTTCTGCG 445  
QY 806 CGATTACCGGCGCACAGAGGAACTCCGATGCTGCTCCGCTTCCGCTGATG 865  
Db 446 CAGCCAGCTGCGCTTGAAGCGCTGCGCTGATGCTTCTCCGCTTCCGCGCG 505  
QY 866 CGCGCATCAGAGAGACATGAGCTTTTGTATGATGATGATGATGATGATGATG 925  
Db 506 CGCGCATCAGAGAGCGGCTGATGCTGATGATGATGATGATGATGATGATGATG 565  
QY 926 AGCGTACCGGAGCGGCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 985  
Db 566 CGCGCTGCGGAGAGAGCATCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 625  
QY 986 GCGCATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1045  
Db 626 GCGTTCCGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 685  
QY 1046 TAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1105  
Db 686 TCGGCTTCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 745  
QY 1106 TCGAGCAGCGAGAGATTTGTAACCTGCTGATGATGATGATGATGATGATG 1165  
Db 746 TCGAGCAGCGAGAGATTTGTAACCTGCTGATGATGATGATGATGATGATGATG 805  
QY 1166 AATTGCGTGCAGCGGTGCGA 1187  
Db 806 AGTTCGCGCGCGCTGCGCA 827

## RESULT 8

US-09-815-242-6828  
; Sequence 6828, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.

APPLICANT: Zykend, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: EITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6828  
LENGTH: 867  
TYPE: DNA  
ORGANISM: Enterococcus faecalis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(867)  
US-09-815-242-6828

Query Match 5.7%; Score 121; DB 9; Length 867;  
Best Local Similarity 53.0%; Pred. No. 8.3e-29;  
Matches 259; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

665 GATGATGCTGTAATTAATGAGAGGCAACCGTGGAGGCGCATCCGGTATATTTTG 744  
328 GATGAGCTGCTTAACGGAGAGAACAAATGAGAGCAAGCTGTTGCAATGGAATC 387  
745 TCCGATTTTCTCTCTGCGGCTCTTTGGGACAGGTCGCGTCCGTCGATCTAGAG 804  
388 ATGAGCGCTAATTTATCATGGGCGATATGGGAGATTTGGTGAATAATCACAGC 447  
805 GCGATTTACCGGCGACAGAGCTGAACTCCCACTGCTGCTCCCTGCTTCCGCTGT 864  
448 TTGTTTGGCGGCGACAGAAAGCATTTACAGTAGATTTTCTACATGCTATGCTGT 507  
865 GCGGCGATGAGAGAGCAATGAGCTTTTGTATGATGATGTCATTAACCGGCTGTG 924  
508 GCCCGTATGAGAGAGAAATTTTTCATTAATGAGAAATGGGAAATTTCCGCGCTTG 567  
925 CAGGCTCAGCGGAGGCGCATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 984  
568 CAACGCGCTAATCAAGAGCGCTTGTATCTTTAGGATTTGATGATCAAGCACTGGC 627  
985 GCGGCGATGAGGCTGCTGCGGCTTCACTGGGCACTCAGCTTTGGGAGAACCGGCGCGAG 1044  
628 GGTGTACCGCAAGTTTTCATGATGATGATGATGATGATGATGATGATGATGAT 687  
1045 ATAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1104  
688 ATCGTTTTCGCTGCGCGCGCTGATTAATGAACAAAGATTCGCAAGATTTGCGAGAT 747  
1105 GTGCGAGCGGAGAGATTTGCTGAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1164  
748 TTTCAAAAGGCGAGATTTTCTTTAGACATGCTTTTGTAGATGATGATGATGATGAT 807  
1165 CAATTCGCT 1173

Db 808 CTTTGGCT 816

RESULT 9  
US-09-070-927A-458  
Sequence 458, Application US/09070927A  
Patent No. US20020120116A1  
GENERAL INFORMATION:  
APPLICANT: Charles A. Kunsch  
Patrick J. Dillon  
Steven Barash  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 982  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/070,927A  
FILING DATE: 04-May-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/046,655  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: 60/044,031  
FILING DATE: 1997-05-06  
APPLICATION NUMBER: 60/066,009  
FILING DATE: 1997-11-14  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB369  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 458:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6021 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 458:  
US-09-070-927A-458

Query Match 5.7%; Score 121; DB 10; Length 6021;  
Best Local Similarity 53.0%; Pred. No. 3.1e-28;  
Matches 259; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

665 GATGATGCTGTAATTAATGAGAGGCAACCGTGGAGGCGCATTCGGTATATTTTG 744  
5321 GATGAGCTGCTTAACGGAGAGAACAAATGAGAGCAAGCTGTTGCAATGGAATC 5380  
745 TCCGATTTTCTCTCTGCGGCTCTTTGGGACAGGTCGCGTCCGTCGATCATAGAG 804  
5381 ATGAGCGCTAATTTATCATGGGCGAGATGAGAAAGATTTGGTGAATAATCACAGC 5440  
805 GCGATTTACCGGCGACAGAGCTGAACTCCCACTGCTGCTCCCTGCTTCCGCTGT 864  
5441 TTGTTTGGCGGCGACAGAAAGCATTTTACAGTAGATTTTCTACATGCTATGCTGT 5500  
865 GCGGCGATGAGAGAGCAATGAGCTTTTGTATGATGATGATGATGATGATGATGAT 924  
5501 GCCGATGAGAGAGAAATTTTTCATGATGATGATGATGATGATGATGATGATGATGAT 5560

QY 925 CAGCGTCAACGCGAGCGCATTTGCCGTTCTGTGTATTTGCGCAATCCACGATGGGT 984  
DB 5561 CAACGGCATTAACAAGCAGCGTTGGTGTATCTTAGGTAATGACATGATCAAGACTGGC 5620  
QY 985 GCGCGCATGCGCTCGTGGGTTTCATCTGGCAATCTCATCTTTGGGAAACCGCGCGCAG 1044  
DB 5621 GGTGTACCGCAAGTTTTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 5680  
QY 1045 ATAGGTTTCCTGGGTCCTGCGTGGGATTAACACGCGGATGCGCTTTCACACGCT 1104  
DB 5661 ATCGGTTTTCCTGGGTCCTGCGTGGGATTAACACGCGGATGCGCTTTCACACGCT 5740  
QY 1105 GTGACGACGCGGAGAAATTTGGTGAATACTGGTGTGATGATGAATTTGTGCGCATC 1164  
DB 5741 TTTAAAGGCGGAGTTTCTTTTAAACATGTTTGTGATGATGATGATGATGATGATGAT 5800  
QY 1165 CAATTCGCT 1173  
DB 5801 CTTTTCGCT 5809

RESULT 10  
US-09-815-242-7288  
; Sequence 7288, Application US/09815242  
; Patent No. US2002061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7288  
; LENGTH: 870  
; TYPE: DNA  
; ORGANISM: Helicobacter pylori  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(870)  
US-09-815-242-7288

Query Match 5.6%; Score 119.2; DB 9; Length 870;  
Best Local Similarity 50.9%; Pred. No. 3.3e-28;  
Matches 283; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 600 TTCTTGAAATGAATCCCAATATGACAACTCAATCAAGCTATGACAGACCTTGA 659  
DB 240 TTACGGCTAATGATCTTTAAATTTGATGATTAAGAGCTTAATAACAACGATTA 299  
QY 660 GCGGCTGGAAGCAAGCCAAATGATGAATCGTAATTACTGGAAGAGGACCGTGA 719

DB 300 AAATATGAAAAGAGATTAACCGCCCAAGCTCAATGATACAGCGGTGAGGTAATCA 359  
QY 720 GGGCATTCGCGTAGCCGTTATTTGTCCGATTTTCTTCTCGCGGCTTTTGGGAC 779  
DB 360 CCGCATGCGCTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 419  
QY 780 GGTGCGGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 839  
DB 420 TGTGAGGAGGCAAAAGATGTAAGAGCAATGATGATGATGATGATGATGATGATGATGAT 479  
QY 840 GGTGCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 899  
DB 480 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 539  
QY 900 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 959  
DB 540 AATGCTAAACGAGCGGCTTTGAAACCGATGATGATGATGATGATGATGATGATGATGAT 599  
QY 960 GTATTTGGCAATCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1019  
DB 600 GCTTTAAGCAATCCACTTAATGAGGCGCTTACGCGATCTTTGCTTTTAAAGGATCT 659  
QY 1020 CACTTTGCGGAACCGCGCGCAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1079  
DB 660 CATTAATCCAGAGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 719  
QY 1080 CACTGGGATGCGCTTCAGACGCTGTCAGACGCGGAGAAATTTGTGAAAATCTGCTGT 1139  
DB 720 TATAGGGCGGATTTGCTGAGGCTTCAACAGCGGAATTTTATAGCATGCTT 779  
QY 1140 GATTGATGAATTTG 1155  
DB 780 GATTGATGAATTTG 795

RESULT 11  
US-09-895-913A-97  
; Sequence 97, Application US/09895913A  
; Patent No. US20020160456A1  
; GENERAL INFORMATION:  
; APPLICANT: Kleantous, Harold  
; APPLICANT: Al-Garawi, Amal  
; APPLICANT: Miller, Charles  
; APPLICANT: Tomb, Jean Francois  
; APPLICANT: Oomen, Raymond P.  
; TITLE OF INVENTION: Identification of Polynucleotides  
; FILE REFERENCE: 06132/043002  
; CURRENT APPLICATION NUMBER: US/09/895,913A  
; PRIOR FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 08/81,227  
; PRIOR FILING DATE: 1997-06-24  
; NUMBER OF SEQ ID NOS: 368  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 97  
; LENGTH: 1053  
; TYPE: DNA  
; ORGANISM: Helicobacter pylori  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (110)...(976)  
US-09-895-913A-97

Query Match 5.6%; Score 119.2; DB 10; Length 1053;  
Best Local Similarity 50.9%; Pred. No. 3.8e-28;  
Matches 283; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 600 TTCTTGAAATGAATCCCAATATGACAACTCAATCAAGCTATGACAGACCTTGA 659  
DB 349 TTACGGCTAATGATCTTTAAATTTGATGATTAAGAGCTTAATAACAACGATTA 408

QY	660	GGGGGCTCGAACCAAGGCCAATATGGATGATCAATCCGTAATTA	CTCGAAGAGGACCGCTGGA	719
Db	409	AAATAACGAAAAAGACTAACCCGCCCAAGCTCAGTATCA	CGGGGAGGCTTAAATCA	468
QY	720	GGGCATTCGGGTAGCCGTTATTTGTGCCATTTTTCCTCTCG	CGCGGTTCTTTGGGCAC	779
Db	469	CCGACATGCTTTGACGATCGTGTGTGTTGATTTTAGCTTTA	TAGGGGGGGAGTTTAGGCTC	528
QY	780	GATGCGCTGCGTGGCGCATATGAAGCCATTACCGCGCCACAG	AGCTGAAACTCCCACT	839
Db	529	TGTGAGAGGCGAAAAGATGTGAAGACATCAATCGCGGGTGC	TTAAAAGAAACGTT	588
QY	840	GCTGCTTCCCTGCTCCGTTCCGATGTGTGCGCGCATGACGA	AGACAATCGACTTTTGTGAT	899
Db	589	ATTGATTTGTTTCAAGGAGTGGGGGGGCTAGATGACAA	AGATCCATTTATGCTATGCA	648
QY	900	GATGTTGCCATTAACCGGGGCTGTGTGAGGTACCGCGAGG	CGCATTTGCCGTTCTGTGT	959
Db	649	AATGCTTAAACGACCGCGGCTTTTAAACCGATGATGAGG	CGCAACTCCCTTTATTTC	708
QY	960	GTATTTCGCATCCACGATGGGTGTGGCCCATGACCTCTG	TGGGGTTTATCTGGGCATCT	1019
Db	709	GCTCTTAAGCATCCCATTTAATGGGGGGGTAGGCCATCT	TTTGTCTTTTAAAGGGATCT	768
QY	1020	CACCTTTTCCGGAACCCGGCGGCGCAGATAGTGTTCCTG	GGTCTCTCGCTGTGTGAGTTAAC	1079
Db	769	CATTATCCAGAGCCACGAGGGCGCATATAGGCTTTTCG	GGGGCTCTAAGGTGATTAACAAC	828
QY	1080	CACCTGGGATGCGCTTCAGACGGTGTGACAGAGCGAGGA	TTTGTGTAACACCTGGAAT	1139
Db	829	TATAGGGCGGATTTGCTCTGAAGGCTTTCAACACAGCGA	ATTTTATTAAGACATGGCTT	888
QY	1140	GATTGATGAATGTG	1155	
Db	889	GATTGATATGATGTG	904	

RESULT 12  
US-09-815-242-9854  
Sequence 9854, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zykkind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OR INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9854  
LENGTH: 993

```

; TYPE: DNA
;; ORGANISM: Salmonella typhimurium
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ... (993)
; US-09-815-242-9854

```

Query Match	5.1%	Score 108;	DB 9;	Length 993;
Best Local Similarity	49.6%	Pred. No. 1.9e-24;		
Matches 276; Conservative	0;	Mismatches 280;	Indels 0;	Gaps 0

QY	638	AAGCGTATGCGAGACCTTGGAGCGGGCTGCAGACGACCAATGCATGATATCGGTA	697
Db	344	AAAATATTAAGACACGACTGGCGTCCGGCGAAGAAACCGCGAGAAACCGGCTGG	403
QY	698	TTACTGAGAAAGCACCGTGGAGGGCAATTCGGTAGCCGTAATTTTGTCCATTTTCT	757
Db	404	TGTCATGAAAGGACCGTTCACGGTATCCGGTGTTCGCGCCGCGTTTGAATTCGGT	463
QY	758	TCCTCGCGGTTCTTTTGGCACGGTCCGCTCGGTGCGCATGAAAGCGATTCCACGG	817
Db	464	TCATGGGCGGCTCATGGGGCTGTGCTGTGGCGACGCTTGCTTCGCGCGTTGAACAG	523
QY	818	CCACAGAGCTAAACTCCCACTGCTGGTCTCCCTGCTTCCGGTGTGTGCGGCATGCAAG	877
Db	524	CGTGGAAAGCAACTGTCCGTTAGTGTCTTCTGTTCGGGCGCGCGGTATGCAAG	583
QY	878	AAGCAATTCGAGCTTTTGTCTATGATGGTGTCCATTAACCGCGGCTGTGACGGTCAACGG	937
Db	584	AAGCGCTGATGTCCCTATGCAATGCGGAAACCTTGCGCGCGCTGTAAATGCAAG	643
QY	938	AAGGCGATTTGCGCGTTCCTGGTGTATTTTGGCGCAATCCACGATGAGTGGCGCAATGCC	997
Db	644	AACGTGTCTGCCCTTAATCTCGGTATTTGACCGATCCGACATGAGCGGCGTTTCCGCCA	703
QY	998	CGTGGGCTATCTGGGCACTTCACTTTTGGGGAACCCGGGCGCGAATAGATTTCCTGG	1057
Db	704	GTTTGTGGATGCTGGGGATCTTAACATGCGCGAGCAAAAGCCCTGATTTGGCTTCGCCG	763
QY	1058	GTCTCTCGCGTGTGAGATTAACCACTGGGCAATGCGGCTTCCAGACGATGCGAGCGG	1117
Db	764	GCCCCGCGTTATGAAACAAACGTTGTGTGAAGACTCCGCCAGAGATTCAGCGGCAAGT	823
QY	1118	AGAAATTTGGTAAAACTGTGTGTATGATGAAATTTGTGTGCGCACTCCAAATTTGCGTGCAG	1177
Db	824	AGTTCTGATGAAAAGGCGGCTTATGATATGATATGATAGTCCGCGTCCGAAATGGCGCTGA	883
QY	1178	CGGTGGCAAAACCCCT	1193
Db	884	AGCTGGGAGACATTCCT	899

RESULT 13  
US-09-815-242-6113  
Sequence 6133, Application US/09615242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlisen, Karl L.  
APPLICANT: Zyekind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA 011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848



```

; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6133
; LENGTH: 915
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(915)
US-09-815-242-6133

```

```

Query Match      4.8%; Score 102.8; DB 9; Length 915;
Best Local Similarity 48.9%; Pred. No. 9.9e-23;
Matches 275; Conservative 0; Mismatches 287; Indels 0; Gaps 0;

```

```

QY 707 AAGGACCGTGGAGGCAATCCGGTAGCCGATTTTTCGATTTTCTTCCCTCGCG 766
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 335 AAGGACCTCTGATGAAGACCGGTTGTCGCGGCAATTCAGTTCCGCTTAAAGGG 394
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 767 GTTCTTTGGGACCGTCCGCTGCGCATCATGAAGCGCATTCACCGCGCACAGAC 826
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 395 GTTCAATGGGGCTCTTTTGGGTGACAGTTTCTGCGCTTACAGAGGGCTGGAAG 454
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 827 TGAACCTCCACTGCTGTCTCCCTGCTTCGCTGCTGCGGCGCATGAGAGAACAATC 886
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 455 ATAACTGCCGCTGATCTGCTTCCGCTCTGCTGCGGCGCATGAGAGAACAATC 514
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 887 GAGCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 946
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 515 TGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 574
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 947 TCCGCTTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1006
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 575 TGCCTGATATCTCCGCTGCTGACCGACCGAGATGGGGGTTTCTGCAAGTTTCGCA 634
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1007 CATCTGGGCACTCACTTTTGGGAAACCGCGCGGCAAGATAGTTTCTGAGTCTGCG 1066
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 635 TGCTGGGCGATCTCAACATCTGTAACCGAAGCGTTAATCGGCTTCCGGTCCGCTG 694
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1067 TGTGAGATTAACCACTGGGATGGCTTCCAGAGGATGATGATGATGATGATGATGATGAT 1126
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 695 TTATCGAAGACAGCGCTTGGCAAAAACGCGCGCTGATTCGAGGCAAGTGAATTCGTA 754
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1127 TGAAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1186
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 755 TCGAAGAAAGGCGCATGACATGATCTGCGCTGCTCGGAATGCGCTGAAACTGGGGA 814
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1187 AAACCTTCAGATTATTCAGCGGAGAGGCAAGATGATGATGATGATGATGATGATGAT 1246
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 815 GCATTCGCGCAAGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 874
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1247 GCGTGGCACTTCGCGGATGTA 1268
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 875 TAGTGTACCCCGGTACCGGA 896
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 14
US-09-815-242-9520
; Sequence 9520, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert

```

```

; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykling, Judith W.
; APPLICANT: Mall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9520
; LENGTH: 867
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(867)
US-09-815-242-9520

```

```

Query Match      4.6%; Score 97.8; DB 9; Length 867;
Best Local Similarity 49.7%; Pred. No. 4.4e-21;
Matches 249; Conservative 0; Mismatches 252; Indels 0; Gaps 0;

```

```

QY 686 ATGAATCGTAACTATGAGAGGACCGGAGGCGATTCGGTAGCCGATTTATTTGT 745
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 332 ATGAAGCGGTGTGACAGAGACTGCTTTATTAAGGCAACATGTGCTTGGGATTA 391
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 746 CCGATTTTCTTCCCTGCGGCTTCTTTGGGACGCTGCGTCCGTCGATCATGAAG 805
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 392 TGAATTTAATTTATATGAGCTTCTATGAGGTGATGAGTAAGTAAGTAAGTAAGTAAG 451
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 806 CGATTCACCGGCGACAGAGCTGAACTCCCACTGCTGCTCCCTGCTCCGCTGCTG 865
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 452 TGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 511
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 866 CGCGCATGACAGAGAGACATGAGCTTTTGTGATGATGATGATGATGATGATGATGATGAT 925
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 512 CCCGTAAGCAGAGAGAGATGAGTCTCATGCAAGATGCTAAGATCTTCCGCGGCTTA 571
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 926 AGCGTACCGGAGCGGATTTGCGCTCTGATGATTTTTCGCAATCCAGATGGGTTG 985
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 572 AAGCTCATTAAGATGCTGCTCTTTTAACGATATTTTACAGATCAAGATCGAGTGG 631
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 986 GCGGCAAGGCTCTGCTGGGTTCACTGCGCATCTCACTTTTGGGAAACCGGCGGCGAGA 1045
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 632 GTGTGACAGCTTTTGTGCTATGAGAGCGGATATCATTTGCTGGAACCAAGAGCTTGG 691
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1046 TAGGTTCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1105
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 692 TTGCTTTTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 751
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1106 TGCAGAGCGGAGAAATTTGTTGTAACCTGCTGATGATGATGATGATGATGATGATGATGAT 1165
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 752 TCCAAAGGCGAGATTCCTATTAAGACATGGCTTTGAGATGCTATTGTCAAAAGAGAG 811
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Oy 1166 ATTGCGTGACGGGTGCAA 1186  
 Db 812 ACTTACCGAGATACGATTGCTA 832

RESULT 15  
 US-09-974-300-5531

; Sequence 5531, Application US/09974300  
 ; Patent No. US20020146721A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Berka, Randy M.  
 ; APPLICANT: Clausen, Ib Groth  
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
 ; FILE REFERENCE: 10085-500-US  
 ; CURRENT APPLICATION NUMBER: US/09/974,300  
 ; CURRENT FILING DATE: 2001-10-05  
 ; PRIOR APPLICATION NUMBER: 09/680,598  
 ; PRIOR FILING DATE: 2000-10-06  
 ; PRIOR APPLICATION NUMBER: 60/279,526  
 ; PRIOR FILING DATE: 2001-03-27  
 ; NUMBER OF SEQ ID NOS: 8481  
 ; SOFTWARE: fastseq for windows Version 4.0  
 ; SEQ ID NO 5531  
 ; LENGTH: 654  
 ; TYPE: DNA  
 ; ORGANISM: Bacillus clausii  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (1)...(654)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-974-300-5531

Query Match 4.6%; Score 96.8; DB 10; Length 654;  
 Best Local Similarity 51.8%; Pred. No. 7.8e-21;  
 Matches 288; Conservative 0; Mismatches 264; Indels 4; Gaps 3;

Oy 636 TCAAGGCTATGAGAGACCTTGGAGCGGCTCGAAGCAAGGCCAATGCGATTCGGT 695  
 Db 84 TCCAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 143  
 Oy 696 AATTACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 754  
 Db 144 GGTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 203  
 Oy 755 CCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 814  
 Db 204 AATTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 263  
 Oy 815 GGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 872  
 Db 264 TGGCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 323  
 Oy 873 GCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 931  
 Db 324 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 383  
 Oy 932 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 991  
 Db 384 TGAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 443  
 Oy 992 TGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1051  
 Db 444 CCGCTAGATTGCGCTCTCTTGGCGATTACAAATTTTGGCGAGCAAAAGACATGATGGCT 503  
 Oy 1052 TCTTGGGTCCTGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1111  
 Db 504 TCGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 563  
 Oy 1112 AGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1171  
 Db 564 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 623

Oy 1172 GTGCAGCGGTGGCAAA 1187  
 Db 624 AAGAAACGCTGACTTA 639

Search completed: November 13, 2003, 06:16:26  
 Job time : 1310.22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 16:48:07 ; Search time 5527.03 Seconds  
(without alignments)  
10902.766 Million cell updates/sec

Title: US-10-024-370-2

Perfect score: 1473

Sequence: 1 gtggaagacgcttcgcac.....gattcagcgttcagcgag 1473

Scoring table:

IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_srb:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pac:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_srs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rsd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1473	100.0	1473	6 AR202320	AR202320 Sequence
2	1473	100.0	1473	6 AX121013	AX121013 Sequence
3	1473	100.0	1473	6 BD004742	BD004742 Replicab1
4	1473	100.0	1473	6 BD163130	BD163130 Novel pol
5	1473	100.0	1575	6 AX066443	AX066443 Sequence
6	1473	100.0	2123	6 CGL17592	Y17592 Corynebacte
7	1473	100.0	2123	6 AR202319	AR202319 Sequence
8	1473	100.0	2123	6 BD004741	BD004741 Replicabl
9	1473	100.0	332050	6 AP005276	AP005276 Corynebac
10	1473	100.0	349980	6 AX127145	AX127145 Sequence
11	602.4	40.9	300750	1 AP005217	AP005217 Corynebac
12	372	25.3	3013	1 CCG117G	X66112 C.gutamicu
13	312.6	21.2	17810	1 AE006979	AE006979 Mycobacte
14	312.6	21.2	37630	1 MTCY31	Z73101 Mycobacteri
15	312.6	21.2	327650	1 BX248337	BX248337 Mycobacte
16	164.4	11.2	342300	1 MLEPRTN8	AL583924 Mycobacte
17	158.4	10.8	10029	1 AE009634	AE009634 Brucella
18	156.8	10.6	10029	1 AE014498	AE014498 Brucella
19	148.6	10.1	260050	1 SME591782	AL591782 Sinorhizo
20	142	9.6	36063	1 MLU15184	U15184 Mycobacteri
21	140.4	9.5	332635	1 AP003005	AP003005 Mesorhizo
22	136.6	9.3	298900	1 AP005937	AP005937 Bradyrhiz
23	132.6	9.0	294250	1 AP001517	AP001517 Bacillus
24	128.4	8.7	7697	1 AE007944	AE007944 Agrobacte
25	128.4	8.7	9508	1 AE008976	AE008976 Agrobacte
26	124.8	8.5	6908	1 SPUS9237	U59237 Synchococc
27	122.8	8.3	4776	1 AB031231	AB031231 Pseudomon
28	122.8	8.3	11976	1 AE004735	AE004735 Pseudomon
29	122.2	8.3	14401	1 AE001518	AE001518 Helicobac
30	121	8.2	304454	1 AE016956	AE016956 Enterococ
31	120.4	8.2	107440	1 AE012365	AE012365 Xanthomon
32	120.2	8.2	311321	1 NMA322491	AL162754 Neisseria
33	119.2	8.1	1053	6 BD092453	BD092453 Identific
34	119.2	8.1	14160	1 AE000604	AE000604 Helicobac
35	119	8.1	3777	1 AF139661	AF139661 Azospiril
36	118.8	8.1	300975	1 AE017013	AE017013 Bacillus
37	118.8	8.1	301200	1 AP005374	AP005374 Thermosyn
38	118.6	8.1	10377	1 AE002422	AE002422 Neisseria
39	118.6	8.1	349980	6 AX044030	AX044030 Sequence
40	117.2	8.0	291804	1 AE017039	AE017039 Bacillus
41	116.6	7.9	13369	1 AE006013	AE006013 Caulobact
42	113.4	7.7	300363	1 AE016781	AE016781 Pseudomon
43	110.8	7.5	14704	1 AE001970	AE001970 Deinococc
44	110.2	7.5	310967	1 AE016869	AE016869 Pseudomon
45	109.8	7.5	15468	1 AE000690	AE000690 Aquifex a

## ALIGNMENTS

RESULT 1  
LOCUS AR202320 1473 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 2 from patent US 6361986.  
ACCESSION AR202320  
VERSION AR202320.1 GI:20256859  
KEYWORDS  
SOURCE  
ORGANISM  
Unknown.  
Unclassified.  
REFERENCE 1 (bases 1 to 1473)  
Tilg.Y., Eikmanns.B., Eggeling.L., Sahn.H. and Mockel.B.  
AUTHORS  
TITLE  
Process for the preparation of L-amino acids by fermentation and  
nucleotide sequences coding for the accDA gene  
JOURNAL  
Patent: US 6361986-A 2 26-MAR-2002;

FEATURES  
source  
BASE COUNT 275 a 414 c 456 g 328 t  
ORIGIN

Query Match 100.0%; Score 1473; DB 6; Length 1473;  
Best Local Similarity 100.0%; Pred. No. 6,7e-293;  
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

1 GTGGAAGCGTTTCCGACTATGTGTGGGCGATGGAACACACTTCAGATTGACGCTC 60
Db 1 GTGGAAGCGTTTCCGACTATGTGTGGGCGATGGAACACACTTCAGATTGACGCTC 60
Qy 61 ATAGACTGGTTTGGACCTTGACAGCTTCATTTCTTGGAATGAACTCCCAATATAC 120
Db 61 ATAGACTGGTTTGGACCTTGACAGCTTCATTTCTTGGAATGAACTCCCAATATAC 120
Qy 121 AACCTCAATCAGGCTATGACAGACCTTGAGCGGGCTCGAAGCAGGCCAAATGCGAT 180
Db 121 AACCTCAATCAGGCTATGACAGACCTTGAGCGGGCTCGAAGCAGGCCAAATGCGAT 180
Qy 181 GAATCGTATATTACTGGAAGAGCAACGTTGAGGGCAATTCCGGTAGCCGTTATTTGTCC 240
Db 181 GAATCGTATATTACTGGAAGAGCAACGTTGAGGGCAATTCCGGTAGCCGTTATTTGTCC 240
Qy 241 GATTTTCTTCTCTGGGGGTTCTTTGGGCAAGGTCGGTCGGTGGCATATGAAGCG 300
Db 241 GATTTTCTTCTCTGGGGGTTCTTTGGGCAAGGTCGGTCGGTGGCATATGAAGCG 300
Qy 301 ATTCAACGGGCGACAGAGCTGAACCTCCACTGCTGTCTCCCTGCTTCCGGTGTGCG 360
Db 301 ATTCAACGGGCGACAGAGCTGAACCTCCACTGCTGTCTCCCTGCTTCCGGTGTGCG 360
Qy 361 CGCATGAGGAAGACATGACCTTTTGTATGATGATGTCTCCATTAACCGCGCTGTGAG 420
Db 361 CGCATGAGGAAGACATGACCTTTTGTATGATGATGTCTCCATTAACCGCGCTGTGAG 420
Qy 421 CGTCAACCGGAGGCGCATTTGCGGTCTCTGGTGTATTTGCGCAATCCACGATGGGTGGC 480
Db 421 CGTCAACCGGAGGCGCATTTGCGGTCTCTGGTGTATTTGCGCAATCCACGATGGGTGGC 480
Qy 481 GCCATGAGGCGTGTGGGGTTCACTGAGGACCTCACTTTTGGGAAACCGGCGCGAGATA 540
Db 481 GCCATGAGGCGTGTGGGGTTCACTGAGGACCTCACTTTTGGGAAACCGGCGCGAGATA 540
Qy 541 GGTTCCTGTGGTCTCTGCGTGTGTGAATTAAACCACTGGGCATGCGCTTTCAGACGATGTG 600
Db 541 GGTTCCTGTGGTCTCTGCGTGTGTGAATTAAACCACTGGGCATGCGCTTTCAGACGATGTG 600
Qy 601 CAGCAGGCGGAGAAATTTGTGTAACCTGTGTGATTTGATGAAATTTGTGCGCACTCCAA 660
Db 601 CAGCAGGCGGAGAAATTTGTGTAACCTGTGTGATTTGATGAAATTTGTGCGCACTCCAA 660
Qy 661 TTGCGGACGCGGTGGCAAAAACCTCAAGTTATTCAACGCGTGAAGSCAACGATCGT 720
Db 661 TTGCGGACGCGGTGGCAAAAACCTCAAGTTATTCAACGCGTGAAGSCAACGATCGT 720
Qy 721 TTTTCTCCAACAACCTCTGCGTGTGGAATTCCGCTGATGAGAGCGAATTCGCGTCTCGT 780
Db 721 TTTTCTCCAACAACCTCTGCGTGTGGAATTCCGCTGATGAGAGCGAATTCGCGTCTCGT 780
Qy 781 GACCCGACAGAGGCTTGGAATCGGGAGATTATGAAAAGTTGGGGGCAACGTCGTCAAG 840
Db 781 GACCCGACAGAGGCTTGGAATCGGGAGATTATGAAAAGTTGGGGGCAACGTCGTCAAG 840
Qy 841 TTTTCTGAGGCGGTGCTGGCGCATTCGCGGCGGTGCGGTGCGTTCCTGGGCGGCATC 900
Db 841 TTTTCTGAGGCGGTGCTGGCGCATTCGCGGCGGTGCGGTGCGTTCCTGGGCGGCATC 900
Qy 901 GGGGGGCGGCGCGTGTGTGATTTGGGAGAGATCGCGCTTCAACGCTTGGGCGGAGAG 960
Db 901 GGGGGGCGGCGCGTGTGTGATTTGGGAGAGATCGCGCTTCAACGCTTGGGCGGAGAG 960

```

Qy 961 CTGGGTTTGGGCGTGTGACATTTTCGTGACGCGGAGTAACCTGCGATGTGTCC 1020  
Db 961 CTGGGTTTGGGCGTGTGACATTTTCGTGACGCGGAGTAACCTGCGATGTGTCC 1020  
Qy 1021 ATCATGACACCTTCGCGCGCGAATTGTGCGAGGCGGCTGAGAGCTCGGCATCGCAAGC 1080  
Db 1021 ATCATGACACCTTCGCGCGCGAATTGTGCGAGGCGGCTGAGAGCTCGGCATCGCAAGC 1080  
Qy 1081 TCATTTGGGCGGACCTTGTTCAGCTTATGAGAGCTTCCTCCCAACCGTTTGCGTCAAT 1140  
Db 1081 TCATTTGGGCGGACCTTGTTCAGCTTATGAGAGCTTCCTCCCAACCGTTTGCGTCAAT 1140  
Qy 1141 ATTGTCAGGCGGTTGGCGGTGACGCGCTGAGGACATGTGCGCCGATCTGTACGCG 1200  
Db 1141 ATTGTCAGGCGGTTGGCGGTGACGCGCTGAGGACATGTGCGCCGATCTGTACGCG 1200  
Qy 1201 GCCGAAAACGCGTGGCTGTCCGATTTGCAACAGAGGCGCTTGCGCATCTCTTCCGC 1260  
Db 1201 GCCGAAAACGCGTGGCTGTCCGATTTGCAACAGAGGCGCTTGCGCATCTCTTCCGC 1260  
Qy 1261 GACACCAACGACGCGCGGAAATCATAGACGACAGGCGGTGAGGCGGACACTTTTA 1320  
Db 1261 GACACCAACGACGCGCGGAAATCATAGACGACAGGCGGTGAGGCGGACACTTTTA 1320  
Qy 1321 AGCCAAAGGCTTATGACAGGAGATGTGCGCGAAACGAGACCTTTGTGAAGAAATCTC 1380  
Db 1321 AGCCAAAGGCTTATGACAGGAGATGTGCGCGAAACGAGACCTTTGTGAAGAAATCTC 1380  
Qy 1381 GGCAACAATCAGCAACGCTCTCTCGAATTGATTAACAATTCGAGAGGCGGACGCGAC 1440  
Db 1381 GGCAACAATCAGCAACGCTCTCTCGAATTGATTAACAATTCGAGAGGCGGACGCGAC 1440  
Qy 1441 AGTGCCTTCAACGATTTGAGCGGTTAGCGCAG 1473  
Db 1441 AGTGCCTTCAACGATTTGAGCGGTTAGCGCAG 1473

RESULT 2  
AX121013  
LOCUS AX121013 1473 bp DNA linear PAT 11-MAY-2001  
DEFINITION Sequence 929 from Patent EP1108790.  
ACCESSION AX121013  
VERSION AX121013.1 GI:14037728  
KEYWORDS  
SOURCE  
ORGANISM  
Corynebacterium glutamicum  
Corynebacterium glutamicum  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
REFERENCE  
1 Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K., Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A., Novel polynucleotides Patent: EP 1108790-A 929 20-JUN-2001;  
JOURNAL KYOMA HAKKO KOGYO CO., LTD. (JP)  
FEATURES  
source  
1. 1473  
/organism="Corynebacterium glutamicum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1718"  
BASE COUNT 275 a 414 c 456 g 328 t  
ORIGIN

Query Match 100.0%; Score 1473; DB 6; Length 1473;  
Best Local Similarity 100.0%; Pred. No. 6,7e-293;  
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

1 GTGGAAGCGTTTCCGACTATGTGTGGGCGATGGAACACACTTCAGATTGACGCTC 60
Db 1 GTGGAAGCGTTTCCGACTATGTGTGGGCGATGGAACACACTTCAGATTGACGCTC 60
Qy 61 ATAGACTGGTTTGGACCTTGACAGCTTCATTTCTTGGAATGAACTCCCAATATAC 120

```

D	b	61	ATAGACTCGGTTTGGACCTGACAGCTTATTTCTTGGATTGAATCCCAATATGAC	120
Q	y	121	AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGGCTGGAAGCAAGCCAAATGCAT	180
D	b	121	AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGGCTGGAAGCAAGCCAAATGCAT	180
Q	y	181	GAATGGGTAAATTAAGTGAAGGACCCGTGAGGGGATTCGGGTACCGTTATTTTGTCC	240
D	b	181	GAATGGGTAAATTAAGTGAAGGACCCGTGAGGGGATTCGGGTACCGTTATTTTGTCC	240
Q	y	241	GATTTTCTCTTCTCGGCGGTTCTTTGGGCAAGGTCGCTCGGTGGCATCATGAAGGCG	300
D	b	241	GATTTTCTCTTCTCGGCGGTTCTTTGGGCAAGGTCGCTCGGTGGCATCATGAAGGCG	300
Q	y	301	ATTCAACCGGCAACAGAGCTGAATCCCACTGCTGTCTCCCTTCCGGTGTGCG	360
D	b	301	ATTCAACCGGCAACAGAGCTGAATCCCACTGCTGTCTCCCTTCCGGTGTGCG	360
Q	y	361	CGCATGCGAAGACAAATGACCTTTTGTCAATGATGATCCATAACCCGCGCTGCGAG	420
D	b	361	CGCATGCGAAGACAAATGACCTTTTGTCAATGATGATCCATAACCCGCGCTGCGAG	420
Q	y	421	CGTCAACCGGCAAGGCGCATTTGCGGTTCCGATGATTTGCGCAATCCCAAGATGGGTGAC	480
D	b	421	CGTCAACCGGCAAGGCGCATTTGCGGTTCCGATGATTTGCGCAATCCCAAGATGGGTGAC	480
Q	y	481	GCCATGCGCTCGTGGGAGTTCAATCTGGGCAATCTCATCTTTTGGGAACCCGCGCGCAGATA	540
D	b	481	GCCATGCGCTCGTGGGAGTTCAATCTGGGCAATCTCATCTTTTGGGAACCCGCGCGCAGATA	540
Q	y	541	GATTTCCTCGGGTCTCGGTGGTGTGAATTAACCACTGGGCAATGCGCTTCCAGACGGTGTG	600
D	b	541	GATTTCCTCGGGTCTCGGTGGTGTGAATTAACCACTGGGCAATGCGCTTCCAGACGGTGTG	600
Q	y	601	CAGCAGGCGGAGAAATTTGGTGAATACTGGTGTGATGATGAATGTGTGCGCACTCCAA	660
D	b	601	CAGCAGGCGGAGAAATTTGGTGAATACTGGTGTGATGATGAATGTGTGCGCACTCCAA	660
Q	y	661	TTGCGTGACGGGTGGCAAAAAACCTCAAGATTATTCAGCCGGTGAAGGCAACGATCGT	720
D	b	661	TTGCGTGACGGGTGGCAAAAAACCTCAAGATTATTCAGCCGGTGAAGGCAACGATCGT	720
Q	y	721	TTTTTCTCCAAACAATCTCTGGCGGTGGCATTTCCGGTGAATGAGGCGAATTCGCGTTCTCGT	780
D	b	721	TTTTTCTCCAAACAATCTCTGGCGGTGGCATTTCCGGTGAATGAGGCGAATTCGCGTTCTCGT	780
Q	y	781	GACCCGACAGGCGCTGGAATCGGGAGATTATGAAACGTTGGGGGCAAGCTCGTCAG	840
D	b	781	GACCCGACAGGCGCTGGAATCGGGAGATTATGAAACGTTGGGGGCAAGCTCGTCAG	840
Q	y	841	CTTTCCTGGTGGCGCTGCTGGCGCATTTGAGCCCGGCGTGTGCGGTTGCCCTGGCGGCAATC	900
D	b	841	CTTTCCTGGTGGCGCTGCTGGCGCATTTGAGCCCGGCGTGTGCGGTTGCCCTGGCGGCAATC	900
Q	y	901	GGGGGCGGGCCGTGTGTCTGATTGGGACAGATCGCCGTTCAAGCTTTGGGCGCAGAG	960
D	b	901	GGGGGCGGGCCGTGTGTCTGATTGGGACAGATCGCCGTTCAAGCTTTGGGCGCAGAG	960
Q	y	961	CTGGGTTTGGCGCGTGTGGGCAATTTGGCGGGGCGGCGAGCTTAAACCTGCGAATGTGTCC	1020
D	b	961	CTGGGTTTGGCGCGTGTGGGCAATTTGGCGGGGCGGCGAGCTTAAACCTGCGAATGTGTCC	1020
Q	y	1021	ATCATGCAACATCTCGGCGCGGAATTTGTGCAAGGCGGCTGAGAGACTCGGCATTCGCAAGC	1080
D	b	1021	ATCATGCAACATCTCGGCGCGGAATTTGTGCAAGGCGGCTGAGAGACTCGGCATTCGCAAGC	1080
Q	y	1081	TGCAATGGCGGCACTTGTCCAAAGTTATCGACGCTCCCTCCCAACCGTTTGGTCAAT	1140
D	b	1081	TGCAATGGCGGCACTTGTCCAAAGTTATCGACGCTCCCTCCCAACCGTTTGGTCAAT	1140
Q	y	1141	ATTGGTCAAGGCGTTGGGCGGTGGCGCGCTGGGCAATGTGCGCCGCGAATCTGGTCTAAGCG	1200
D	b	1141	ATTGGTCAAGGCGTTGGGCGGTGGCGCGCTGGGCAATGTGCGCCGCGAATCTGGTCTAAGCG	1200

OY	1201	GCCGAAAACGGTGGCTGTGCGAATTGGCAACAAGGAGCGCTCGGCATCTTTCGC	1260
Db	1201	GCCGAAAACGGTGGCTGTGCGAATTGGCAACAAGGAGCGCTCGGCATCTTTCGC	1260
OY	1261	GACACCAACGCGCGGAAATCATAGAGGACAAAGCGGTGACGCGCACCTTTTA	1320
Db	1261	GACACCAACGCGCGGAAATCATAGAGGACAAAGCGGTGACGCGCACCTTTTA	1320
OY	1321	AGCCAAAGGCTTATCGACGGGATGTGGCGCGAAACCGAGCACTTTGTTGAAGAAATCTC	1380
Db	1321	AGCCAAAGGCTTATCGACGGGATGTGGCGCGAAACCGAGCACTTTGTTGAAGAAATCTC	1380
OY	1381	GGCAACATCGACAAGCGCTTCTCCGAATTGGATTACATTCGAGAGAGCGCGGAC	1440
Db	1381	GGCAACATCGACAAGCGCTTCTCCGAATTGGATTACATTCGAGAGAGCGCGGAC	1440
OY	1441	AGTGGCTTCAAGATTGAAGCGTTTGAAGCGAG	1473
Db	1441	AGTGGCTTCAAGATTGAAGCGTTTGAAGCGAG	1473

RESULT 3	BD004742	1473 bp	DNA	linear	PAT 31-JAN-2002
LOCUS	BD004742				
DEFINITION	Replicable DNA, amino acid sequence, <i>Corynebacterium microorganism</i> , shuttle vector, and process for producing L-amino acid.				
ACCESSION	BD004742				
VERSION	BD004742.1	GI:18632703			
KEYWORDS	JP 2001008693-A/2.				
SOURCE	<i>Corynebacterium glutamicum</i>				
ORGANISM	<i>Corynebacterium glutamicum</i>				

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 1473)	Iljke,I., Eggering,L., Rickmans,B., Zamu,H. and Meck,V.	Replicable DNA, amino acid sequence, Corynebacterium microorganism, shuttle vector, and process for producing L-amino acid	Patent: JP 2001008693-A 2 16-0AN-2001;
DEUSSA HUELS AG,FORSCHUNGSZENTRUM JUELICH GMBH	OS	Corynebacterium glutamicum	

	FN	UF	ZOV,U06893-N/2	
	PD	16-JAN-2001		
	PF	24-MAY-2000	JF 2000153547	
	PR	27-MAY-1999	DE 19924365.4	
	PI	VEENNE TIJKE, LOTHAR EGGERING, BERNHARD EICKMANS, HERRMANN ZAMU,		
	PL	VETLYNA MECKEL		
	PC	C12N15/09, C12N1/21, C12P13/04, C12P13/06, C12P13/08, PC		
		C12P13/08		
	PC	C12P13/12, C12P13/20, C12P21/02// (C12N15/09, C12R1:15), (C12N1/21,		
		C12R1:15),		
	PC	(C12P13/04, C12R1:15), (C12P13/06, C12R1:15), (C12P13/08, C12R1:15), PC		
		(C12P13/12, C12R1:15), (C12P13/20, C12R1:15), (C12P21/02, C12R1:15), PC		
		C12N15/00		
	PC	(C12N15/00, C12R1:15)		
	CC			
	FT			
FEATURES	Key	Location/Qualifiers		
	CDS	(1). (1473).		
source	Location/Qualifiers			
	1. .1473			
	/organism="Corynebacterium glutamicum"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:1718"			
BASE COUNT	275 a	414 c	456 g	328 t
ORIGIN				

Query Match	100.0%	Score 1473	DB 6	Length 1473
Best Local Similarity	100.0%	Pred. No. 6.7e-293		
Matches 1473	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	GTGGAGAAAGCGTTTCCGACTATGCTGTGTGGGCGATGGAAACACCTTCAGCATTGACGCTC	60	

Db 1 GTGAGAGCGTTTCCGACTATGCTGAGGAGCATGAAACACATTCAGATGACGCTC 60  
Qy 61 ATAGACTGGTTTGGACCTTGAACGCTTCAATTTCTTGGAAATGAAATCTCCCAATATGAC 120  
Db 61 ATAGACTGGTTTGGACCTTGAACGCTTCAATTTCTTGGAAATGAAATCTCCCAATATGAC 120  
Qy 121 AACCTCAATCAAGGCTATGACAGAGACTTGGAGCGGGCTCGAAGCAAGCCCAATGCGAT 180  
Db 121 AACCTCAATCAAGGCTATGACAGAGACTTGGAGCGGGCTCGAAGCAAGCCCAATGCGAT 180  
Qy 181 GAATGGTAATTAATCTGAGAGAGCAACCGTGAAGGCAATTCGGATGACCGTTATTTGTTC 240  
Db 181 GAATGGTAATTAATCTGAGAGAGCAACCGTGAAGGCAATTCGGATGACCGTTATTTGTTC 240  
Qy 241 GATTTTCTTCTCCGCGGGTCTTGGAGCAACCGTGGTGGTGGTGGTGGTGGTGGTGGTGG 300  
Db 241 GATTTTCTTCTCCGCGGGTCTTGGAGCAACCGTGGTGGTGGTGGTGGTGGTGGTGGTGG 300  
Qy 301 ATTCAACCGGCGACAGAGCTGAACTCCCACTGCTGTCTCCCTGCTTCCGATGATGAG 360  
Db 301 ATTCAACCGGCGACAGAGCTGAACTCCCACTGCTGTCTCCCTGCTTCCGATGATGAG 360  
Qy 361 CGCATGAGAGAGCAATCGAGCTTTGTATGATGATGATGATGATGATGATGATGATGATG 420  
Db 361 CGCATGAGAGAGCAATCGAGCTTTGTATGATGATGATGATGATGATGATGATGATGATG 420  
Qy 421 CGTACCGGCGAGGCGCAATTCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
Db 421 CGTACCGGCGAGGCGCAATTCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
Qy 481 GCCATGAGCTGCTGAGGCTTCACTGAGGATCTCACTTTTGGAGAACCGGCGCGAGATA 540  
Db 481 GCCATGAGCTGCTGAGGCTTCACTGAGGATCTCACTTTTGGAGAACCGGCGCGAGATA 540  
Qy 541 GATTTCTGAGGCTCTGCGTGGTGGATTAACCATGCGCATGCGCTTCCAGACGCTGTG 600  
Db 541 GATTTCTGAGGCTCTGCGTGGTGGATTAACCATGCGCATGCGCTTCCAGACGCTGTG 600  
Qy 601 CAGAGGCGGAGAAATTTGATGAAATCTGATGATGATGATGATGATGATGATGATGATG 660  
Db 601 CAGAGGCGGAGAAATTTGATGAAATCTGATGATGATGATGATGATGATGATGATGATG 660  
Qy 661 TTGCGTGAAGCGGTGGCAAAAACCTCAAGGTTATTCAGCGGTAGAGCAACGATGCT 720  
Db 661 TTGCGTGAAGCGGTGGCAAAAACCTCAAGGTTATTCAGCGGTAGAGCAACGATGCT 720  
Qy 721 TTTTCTCAACAATCTCTGCGTGGCACTTCCGATGATGAGGCGATGCGGTTCTGCT 780  
Db 721 TTTTCTCAACAATCTCTGCGTGGCACTTCCGATGATGAGGCGATGCGGTTCTGCT 780  
Qy 781 GACCGGAGAGGCTGGAATCGGGAGATTAATGAAAGCTTGGGGGAGAGAGCTGCTCAAG 840  
Db 781 GACCGGAGAGGCTGGAATCGGGAGATTAATGAAAGCTTGGGGGAGAGAGCTGCTCAAG 840  
Qy 841 CTTTCTGAGTGGCGTGGCGCTGAGCGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCG 900  
Db 841 CTTTCTGAGTGGCGTGGCGCTGAGCGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCG 900  
Qy 901 GGGGGCGGCGCGCTGAGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Db 901 GGGGGCGGCGCGCTGAGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Qy 961 CTTGCGTTTGGCGGCTGAGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Db 961 CTTGCGTTTGGCGGCTGAGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Qy 1021 ATCATGAGACCTCGGCGCGCGAATTTGTCAGAGCGGCTGAGAGAGCTGAGAGCTGAGAG 1080  
Db 1021 ATCATGAGACCTCGGCGCGCGAATTTGTCAGAGCGGCTGAGAGAGCTGAGAGCTGAGAG 1080  
Qy 1081 TCGATTTGGCGGACCTTGTCCAAAGCTTAATGAGAGCTTCCCTCCCAAGCTTTGGTCAAT 1140  
Db 1081 TCGATTTGGCGGACCTTGTCCAAAGCTTAATGAGAGCTTCCCTCCCAAGCTTTGGTCAAT 1140

Qy 1141 ATTGTCAGGGCGTTTGGCGGATGGCGGCTGAGCCATGCTGACCGGCAATCTGATACGCG 1200  
Db 1141 ATTGTCAGGGCGTTTGGCGGATGGCGGCTGAGCCATGCTGACCGGCAATCTGATACGCG 1200  
Qy 1201 GCCGAAACAGCGTGGCTGTCGATGTCAGACAGAGAGGCGCTTGGCAATCTTTCGCG 1260  
Db 1201 GCCGAAACAGCGTGGCTGTCGATGTCAGACAGAGAGGCGCTTGGCAATCTTTCGCG 1260  
Qy 1261 GACACCAACAGCGCGGAGAAATCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
Db 1261 GACACCAACAGCGCGGAGAAATCATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
Qy 1321 AGCCAGAGGCTTATGAGAGGATGTCGCGGAGAAACGAGGACATTTGTAAGAAATCTC 1380  
Db 1321 AGCCAGAGGCTTATGAGAGGATGTCGCGGAGAAACGAGGACATTTGTAAGAAATCTC 1380  
Qy 1381 GGCAATATGAGCAACGCGCTTCTCGAATTTGATTAACATTCGAGAGAGGCGGAGCGGAC 1440  
Db 1381 GGCAATATGAGCAACGCGCTTCTCGAATTTGATTAACATTCGAGAGAGGCGGAGCGGAC 1440  
Qy 1441 AGTGGCTTACAGCATTTGAGCGCTTTAGCGCAG 1473  
Db 1441 AGTGGCTTACAGCATTTGAGCGCTTTAGCGCAG 1473  
  
RESULT 4  
BD163130  
LOCUS BD163130 1473 bp DNA linear PAT 17-JAN-2003  
DEFINITION Novel polynucleotide.  
ACCESSION BD163130  
VERSION BD163130.1 GI:27868892  
KEYWORDS JP 2002191370-A/929.  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 1473)  
AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K., Yokoi, H., Tateishi, N., Senoo, A., Ikeda, M. and Ozaki, A.  
TITLE Novel polynucleotide  
JOURNAL Patent: JP 2002191370-A 929 09-JUL-2002;  
COMMENT KYOMA HAKKO KOGYO CO LTD  
OS Corynebacterium glutamicum  
PN JP 2002191370-A/929  
PD 09-JUL-2002  
PF 15-DEC-2000 JP 2000405096  
PI SATOSHI NAKAGAWA, HIROSHI MIZOGUCHI, SEIKO ANDO, MIKIO HAYASHI, KEIRO OCHIAI,  
PI HARUHIKO YOKOI, NAKO TATEISHI, AKIHIRO SENOO, MASATO IKEDA, AKIO OZAKI  
PI  
PC C12N1/09, C12N1/05, C07K14/34, C07K16/12, C07K16/40, C12M1/00, PC C12N1/15,  
PC C12N1/19, C12N1/21, C12N5/10, C12N9/00, C12N9/02, C12P1/40, C12P13/04, C12P13/08,  
PC C12P19/00, C12P19/34, C12P21/02, C12P1/37, C12Q1/68, G01N33/53, PC G01N33/56,  
PC G01N33/569, G01N33/68, G01N37/00, C12P21/08, C12N1/21, C12R1/15, PC C12N1/21, C12R1/13, C12N1/21, C12R1/01, C12P13/08, C12R1/15,  
PC C12N1/00, C12N1/00  
CC Novel polynucleotide  
FH Key  
FT source 1.1473  
location/Qualifiers  
1.1473  
/organism='Corynebacterium glutamicum'.  
location/Qualifiers  
1.1473  
/organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'  
BASE COUNT 275 a 414 c 456 g 328 t  
ORIGIN  
Query Match 100.0%; Score 1473; DB 6; Length 1473;



Best Local Similarity 100.0%; Pred. No. 6.7e-293;  
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTGGAGAGAGGCTTTCCGACTATGTGTGGGGGCTGAGACACATTCAGATTCAGCTC 60
Db 1 GTGGAGAGAGGCTTTCCGACTATGTGTGGGGGCTGAGACACATTCAGATTCAGCTC 60
QY 61 ATAGACTCGGTTTGAACCTGACAGCTTCACTTTGATGAACTCCCAATATGAC 120
Db 61 ATAGACTCGGTTTGAACCTGACAGCTTCACTTTGATGAACTCCCAATATGAC 120
QY 121 AACCTCAATGAAGGCTATGACAGACCTTGGAGGGGCTGAGAGCAAGGCCAAATGCA 180
Db 121 AACCTCAATGAAGGCTATGACAGACCTTGGAGGGGCTGAGAGCAAGGCCAAATGCA 180
QY 181 GAATCGGTATTAATGAGAGAGGACCGTGAAGGGGCTTCGGGTAGCCGTTATTTGTC 240
Db 181 GAATCGGTATTAATGAGAGAGGACCGTGAAGGGGCTTCGGGTAGCCGTTATTTGTC 240
QY 241 GATTTTCTTCTCCGCGGTTCTTTGGGACGAGTCCGTCGGTGCGCATATGAAGGC 300
Db 241 GATTTTCTTCTCCGCGGTTCTTTGGGACGAGTCCGTCGGTGCGCATATGAAGGC 300
QY 301 ATTACCGGCGCACAAGAGCTGAACTCCCACTGCTGCTCCCTGCTTCGGTGTGCG 360
Db 301 ATTACCGGCGCACAAGAGCTGAACTCCCACTGCTGCTCCCTGCTTCGGTGTGCG 360
QY 361 CGCATGAGAGAGACAAATGAGGCTTTTGTATGATGATGATGATGATGATGATGATG 420
Db 361 CGCATGAGAGAGACAAATGAGGCTTTTGTATGATGATGATGATGATGATGATGATG 420
QY 421 CGTCAACGCGAGGCGCATTTGCGTCTGCTGATTTTGGGCAATCCAGATGGGAGG 480
Db 421 CGTCAACGCGAGGCGCATTTGCGTCTGCTGATTTTGGGCAATCCAGATGGGAGG 480
QY 481 GCCATGCGCTCGTGGGCTTCACTGAGGCACTTCATTTTGGGCAATCCAGATGGGAGG 540
Db 481 GCCATGCGCTCGTGGGCTTCACTGAGGCACTTCATTTTGGGCAATCCAGATGGGAGG 540
QY 541 GGTTCCTGGGCTCGTGGGCTTCACTGAGGCACTTCATTTTGGGCAATCCAGATGGGAGG 600
Db 541 GGTTCCTGGGCTCGTGGGCTTCACTGAGGCACTTCATTTTGGGCAATCCAGATGGGAGG 600
QY 601 CAGCAGCGCGAGATTTGGTGAATACTGGTGTGATGATGATGATGATGATGATGATG 660
Db 601 CAGCAGCGCGAGATTTGGTGAATACTGGTGTGATGATGATGATGATGATGATGATG 660
QY 661 TTGCGTGACGCGGTGCAAAAACCTCAAGGTTTATTCAGCCGTTAGAGGCAACGATCGT 720
Db 661 TTGCGTGACGCGGTGCAAAAACCTCAAGGTTTATTCAGCCGTTAGAGGCAACGATCGT 720
QY 721 TTTTCCCAACAACCTCGGCGTGGGCACTTCGGGTATGAGAGGCAATTCGGGTTTCGT 780
Db 721 TTTTCCCAACAACCTCGGCGTGGGCACTTCGGGTATGAGAGGCAATTCGGGTTTCGT 780
QY 781 GACCCGAGAGGCTGGAATCGGAGGATTTATGAGAAACCTGGGGGAGAGAGGCTGTCAG 840
Db 781 GACCCGAGAGGCTGGAATCGGAGGATTTATGAGAAACCTGGGGGAGAGAGGCTGTCAG 840
QY 841 CTTTTCGTTGCGGCTGCGGCAATTCAGCCGCTGTGCGGCTGCGGCTGCGGCAATC 900
Db 841 CTTTTCGTTGCGGCTGCGGCAATTCAGCCGCTGTGCGGCTGCGGCTGCGGCAATC 900
QY 901 GGGGGCGGGGCGGCTGCTGATTTGGGAGAGATTCGCGGCTTCAACCTTGGGGCGGAGAG 960
Db 901 GGGGGCGGGGCGGCTGCTGATTTGGGAGAGATTCGCGGCTTCAACCTTGGGGCGGAGAG 960
QY 961 CTGCGTTTTCGCGCTGCTGATTTTCGCTGCGGCGGAGCTTAACTTCCGATCGGTGTC 1020
Db 961 CTGCGTTTTCGCGCTGCTGATTTTCGCTGCGGCGGAGCTTAACTTCCGATCGGTGTC 1020
QY 1021 ATCATGACACTTCGGGCGGCAATTTGTCGAGGCGGCTGAGAGCTTGGGCAATCGCAAGC 1080
Db 1021 ATCATGACACTTCGGGCGGCAATTTGTCGAGGCGGCTGAGAGCTTGGGCAATCGCAAGC 1080

```

```

Db 1021 ATCATGACACTTCGGGCGGCAATTTGTCGAGGCGGCTGAGAGCTTGGGCAATCGCAAGC 1080
QY 1081 TCGATTGCGGCGACCTTTGTCAAGCTTATGAGGCTCCCTCCCAACCTTTGGTCAAT 1140
Db 1081 TCGATTGCGGCGACCTTTGTCAAGCTTATGAGGCTCCCTCCCAACCTTTGGTCAAT 1140
QY 1141 ATTGTCAGGGGCTTTGGGCGGCTGCGGCTGAGGCAATCGGCTGAGGCTGAGGCT 1200
Db 1141 ATTGTCAGGGGCTTTGGGCGGCTGCGGCTGAGGCAATCGGCTGAGGCTGAGGCT 1200
QY 1201 GCGGAAACGCGTGTGCTGCTGATTCGCAATGCAACGAGGCGGCTCGGCAATCGTTCGCG 1260
Db 1201 GCGGAAACGCGTGTGCTGCTGATTCGCAATGCAACGAGGCGGCTCGGCAATCGTTCGCG 1260
QY 1261 GACACCAACGACCGCGGGAATATGAGGCAACGAGGCTGAGGCGGCAATCTTTA 1320
Db 1261 GACACCAACGACCGCGGGAATATGAGGCAACGAGGCTGAGGCGGCAATCTTTA 1320
QY 1321 AGCCAAAGGCTTATGACGAGGATTCGCGGCAACGAGCACTTTGTAAGAAATTC 1380
Db 1321 AGCCAAAGGCTTATGACGAGGATTCGCGGCAACGAGCACTTTGTAAGAAATTC 1380
QY 1381 GGCACATGACGACGCGCTCTCCGATTTGATTAACATTCGAGAGGCGGAGCGGAC 1440
Db 1381 GGCACATGACGACGCGCTCTCCGATTTGATTAACATTCGAGAGGCGGAGCGGAC 1440
QY 1441 AGTCGCTTCAACGATTTGAGGCTTTAGCGGAG 1473
Db 1441 AGTCGCTTCAACGATTTGAGGCTTTAGCGGAG 1473

```

RESULT 5  
AX066443  
LOCUS  
DEFINITION  
Sequence 25 from Patent WO010805.  
AX066443  
VERSION  
AX066443.1 GI:12544151  
KEYWORDS

SOURCE  
ORGANISM  
Corynebacterium glutamicum  
Corynebacterium glutamicum  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacteriaceae; Corynebacterium.

REFERENCE  
AUTHORS  
Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Habermeyer, G.  
TITLE  
corynebacterium glutamicum genes encoding proteins involved in  
membrane synthesis and membrane transport  
JOURNAL  
BASF AKTIENGESELLSCHAFT (DE)  
FEATURES  
source  
1.1575  
Location/Qualifiers  
/organism="Corynebacterium glutamicum"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1718"  
101..1555  
/note="RXN02173"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CA26449.1"  
/db\_xref="GI:12544151"  
/translation="NMVMEHTSALTLIDSVLPDSFISWNETPOYDNIONGYAEITL  
RARSKACDESVITGEGVEGIPVAIVLSDSFGSGISGTVASVPIMKAIHRAETIKL  
PLVSPASGARMEDNRAFVMMVMSITAAVORHREAHPLVLYLRNPTMGAMASWGS  
SGHLTPAEPCMOIEFLGRVVELITGHALPVGVOAENLVKTVIGTIVSPLOASAAV  
AKTLKIVQPEVATDRFSTTGGVALPVWEALARSDDPQRPQIGETIMETIGADVLSG  
ARAGALSVAVALARIGRPVVLIGDPRFTLPQELRPARGISLARBNLPVSI  
IDTGAELSOAAEELGASJARTLSKIDAPLPVSVIIGGVGGALAMPDLVY  
AAENAMSLPPEGASALIPDTYHAEILIRROCVQMAHLLSGLIDIVAEHFIVE  
EILDTISMALESNNPERARDRPFTRFETLQ"

BASE COUNT  
ORIGIN  
309 a 434 c 474 g 358 t

Query Match  
Best Local Similarity 100.0%; Score 1473; DB 6; Length 1575;  
Best Local Similarity 100.0%; Pred. No. 6.6e-293;

	Matches	1473:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:
OY	1	GTGGAGAAAGCGTTTTC	CGACTATATGGTGTGGGGCATGAA	CACACTTCAGCATTTGACGCTC	60					
Db	80	GTGGAGAAAGCGTTTTC	CGACTATATGGTGTGGGGCATGAA	CACACTTCAGCATTTGACGCTC	139					
OY	61	ATGACCTGGTTTGGACCCCTG	ACAGGCTTCATTTCTGTGAATGAA	CACTCCCAATATGAC	120					
Db	140	ATGACCTGGTTTGGACCCCTG	ACAGGCTTCATTTCTGTGAATGAA	CACTCCCAATATGAC	199					
OY	121	AACCTCAATCAAGGCTATG	ACAGAGACTTTGAGCGGGCTCGA	AGCAAGGCCAAATGCGAT	180					
Db	200	AACCTCAATCAAGGCTATG	ACAGAGACTTTGAGCGGGCTCGA	AGCAAGGCCAAATGCGAT	259					
OY	181	GAATCGTAATTTACTGGA	AGAGGACCCGTGAGGGCAATTCGGTAG	ACCGTTATTTTGTCC	240					
Db	260	GAATCGTAATTTACTGGA	AGAGGACCCGTGAGGGCAATTCGGTAG	ACCGTTATTTTGTCC	319					
OY	241	GATTTTCTCTTCCTGCGGCTT	CTTTGGGACAGGTCGGTGGGTGGGATCAT	TGAAGGCG	300					
Db	320	GATTTTCTCTTCCTGCGGCTT	CTTTGGGACAGGTCGGTGGGTGGGATCAT	TGAAGGCG	379					
OY	301	ATTACCGCGCCACAGAGCT	GAACTCCCACTGCTGATCTCCCTGCTTC	CGAGTGTGCG	360					
Db	380	ATTACCGCGCCACAGAGCT	GAACTCCCACTGCTGATCTCCCTGCTTC	CGAGTGTGCG	439					
OY	361	CGCATGACGAGAGACAAAT	CGACTTTTGTCAATGATGTGTCCATA	ACCGCGGCTGTGAG	420					
Db	440	CGCATGACGAGAGACAAAT	CGACTTTTGTCAATGATGTGTCCATA	ACCGCGGCTGTGAG	499					
OY	421	CGTACCCGCGAGGGCCAT	TTTGGCCGTTCTGTGTATTTTGC	GAAATCCACGATGGGTGGC	480					
Db	500	CGTACCCGCGAGGGCCAT	TTTGGCCGTTCTGTGTATTTTGC	GAAATCCACGATGGGTGGC	559					
OY	481	GCCATAGGCGCTCGTGGG	ATTCACTGAGGCACTCATTTTGC	AGCAACCGGACGACGATA	540					
Db	560	GCCATAGGCGCTCGTGGG	ATTCACTGAGGCACTCATTTTGC	AGCAACCGGACGACGATA	619					
OY	541	GCTTTCCTGGGTCCTCG	CGGTGGAGTTAACTACTGTGGCATGCGCTT	CCAGACGATGTG	600					
Db	620	GCTTTCCTGGGTCCTCG	CGGTGGAGTTAACTACTGTGGCATGCGCTT	CCAGACGATGTG	679					
OY	601	CAGCAGGGCGGAGAAAT	TGTGTGAAACTGTGTGATATGAA	ATTGTGTCCGCACTCCAA	660					
Db	680	CAGCAGGGCGGAGAAAT	TGTGTGAAACTGTGTGATATGAA	ATTGTGTCCGCACTCCAA	739					
OY	661	TTGCGTGCAGCGGTGSC	CAAAAACCCCTCAAGGTTATTCAGCGGTG	AGGCAACCGATCGT	720					
Db	740	TTGCGTGCAGCGGTGSC	CAAAAACCCCTCAAGGTTATTCAGCGGTG	AGGCAACCGATCGT	799					
OY	721	TTTTCTTCAACAACATCT	CTGTGCGTGGCATCTCCGATGATG	AGGCGATTTGCGGCTCTCGT	780					
Db	800	TTTTCTTCAACAACATCT	CTGTGCGTGGCATCTCCGATGATG	AGGCGATTTGCGGCTCTCGT	859					
OY	781	GACCCGCAAGAGGCTTGA	ATCGGGAGATTTATGAAA	CGTTGGGGGCAAGCTGTCTAAG	840					
Db	860	GACCCGCAAGAGGCTTGA	ATCGGGAGATTTATGAAA	CGTTGGGGGCAAGCTGTCTAAG	919					
OY	841	CTTTCTGAGTGGCGGTG	CGGTCGCAATTTAGCCCGGCTGTGGCGGCTTG	ACCTGGCGGGGCAATC	900					
Db	920	CTTTCTGAGTGGCGGTG	CGGTCGCAATTTAGCCCGGCTGTGGCGGCTTG	ACCTGGCGGGGCAATC	979					
OY	901	GAGGAGCCGAGCGCTGT	GTGATTTGGGACAGATTCGCGCTT	CACGCTTTGGGCGGACGAG	960					
Db	980	GAGGAGCCGAGCGCTGT	GTGATTTGGGACAGATTCGCGCTT	CACGCTTTGGGCGGACGAG	1033					
OY	961	CTGCGTTTTCGCGCTGT	GGGCAATTTGCTGTGGCGGCGGCACTTAA	CCGTGCCATGTGTCC	1020					
Db	1040	CTGCGTTTTCGCGCTGT	GGGCAATTTGCTGTGGCGGCGGCACTTAA	CCGTGCCATGTGTCC	1099					
OY	1021	ATCATTCGACACTCCGG	GGCGCGAATTTGTGACAGGGCGCTG	AGAGACTGTGGCATTCGAAGC	1087					
Db	1100	ATCATTCGACACTCCGG	GGCGCGAATTTGTGACAGGGCGCTG	AGAGACTGTGGCATTCGAAGC	1155					

QY	1061	TCGATTTGGCGGCACCTTGTCCAAACCTTAATGAGAGCTCCCTCCCGCACCGTTTGGTGCAAT	1140		
Db	1160	TCGATTTGGCGGCACCTTGTCCAAACCTTAATGAGAGCTCCCTCCCGCACCGTTTGGTGCAAT	1219		
QY	1141	ATTGTGACGAGCGCTGTGGCGGTGTGGCGCGCTGTGGCCATGCTGCCCCGATCTGTGTACGCG	1200		
Db	1220	ATTGTGACGAGCGCTGTGGCGGTGTGGCGCGCTGTGGCCATGCTGCCCCGATCTGTGTACGCG	1279		
QY	1201	GCCGAAAACGCTGTGGCTGTCCGCAATTGCGACACGAGGGCGCTGTGGCATCTTTCGCCG	1260		
Db	1280	GCCGAAAACGCTGTGGCTGTCCGCAATTGCGACACGAGGGCGCTGTGGCATCTTTCGCCG	1339		
QY	1261	GACACCAACACGCGCGGGGAAATATATAGAGGACAGAGGCGTGCAGGCGACGCACTTTTA	1320		
Db	1340	GACACCAACACGCGCGGGGAAATATATAGAGGACAGAGGCGTGCAGGCGACGCACTTTTA	1399		
QY	1321	AGCCAGAGGCGCTTATCGACGGGATCGTGCAGAACCGAGCACTTTGTGTAAGAAATCTC	1360		
Db	1400	AGCCAGAGGCGCTTATCGACGGGATCGTGCAGAACCGAGCACTTTGTGTAAGAAATCTC	1459		
QY	1361	GGCACATCAGCAACGCCCTTCCGAAATTGGATTAACAATCCGAGAGAGGGCGGACGGAC	1440		
Db	1460	GGCACATCAGCAACGCCCTTCCGAAATTGGATTAACAATCCGAGAGAGGGCGGACGGAC	1519		
QY	1441	AGTCGCTTCAACAGATTGAGCGTTTGGCGCAG	1473		
Db	1520	AGTCGCTTCAACAGATTGAGCGTTTGGCGCAG	1552		
RESULT 6	CG117592	2123 bp	DNA	linear	BCT 26-JUN-2001
LOCUS	CG117592				
DEFINITION	Corynebacterium glutamicum accDA gene.				
ACCESSION	Y17592				
VERSION	Y17592.1	GI:14572581			
KEYWORDS	accDA gene; acetyl-CoA carboxylase; carboxyltransferase.				
SOURCE	Corynebacterium glutamicum				
ORGANISM	Corynebacterium glutamicum				
	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriineae; Corynebacteriaceae; Corynebacterium.				
REFERENCE	1	Kim, K., Peters-Wendisch, P. G., Sahm, H. and Eikmanns, B. J.			
AUTHORS	Unpublished				
REFERENCE	2	(bases 1 to 2123)			
AUTHORS	Peters-Wendisch, P.G.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-JUN-1998) P.G. Peters-Wendisch, University of California at Berkeley, Dept. Plant and Microbial Biology, Kustu Lab, 111 Koshland Hall, Berkeley CA 94720, USA				
COMMENT	Related sequence x66112.				
FEATURES	location/qualifiers				
source	1..2123				
	/organism="Corynebacterium glutamicum"				
	/mol_type="genomic DNA"				
	/strain="ATCC 13032"				
	/db_xref="taxon:1718"				
	508..1983				
	/gene="accDA"				
	508..1983				
	/gene="accDA"				
	/codon_start=1				
	/transl_table=1				
	/product="putative carboxyltransferase subunit of acetyl-CoA carboxylase"				
	acetyl-CoA carboxylase				
	/protein_id="CAC42827.1"				
	/db_xref="GI:14572582"				
	/translation="MEKRPFTMWGMEHTSALITLISVLPDPSISWNETPOYDNLNQGYATLIRARSKACDSVITGEGTGGIPIVAILSDPSFLGSLGVASRIKALIHRAITLKIPLVSPASGARGMOENRPAVMVSIITAAQORREAHLPVLYLRPTMGGAAMASWSSGHLTFABEQAQIGFLGPRVELITGHALPDGYOQENLYKTEVIDGIVPDLQAAAVKTLKVIQPEATIDRESPTTPGVALPVEWALIASRDFORFGIGIMETLDAIVTLGASRAGALSPAVVALARIGRPPVVLIGDRRFTLGPELRFARGLSIARL				
gene					
CDS					

NUPVSIIDTSGAELISQAELIAGSIARTLSKIDAPLPVSVIIIGGVGALAM  
 LPADIVYAENMWSALPEEGASALIFRDTNHAABIIIEOGVQAHALLSQGLIDIVA  
 ETEHVEBIEITGISNALSSELDNNPERAGDSRPTREPRLAQ"

BASE COUNT 460 a 574 c 592 g 497 t  
 ORIGIN

Query Match 100.0%; Score 1473; DB 1; Length 2123;  
 Best Local Similarity 100.0%; Pred. No. 6,4e-293;  
 Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTGAGAACGCTTTTCCGATGATGTTGGGCGCATGAAACACCTTCAGATTGACGCTC 60  
 508 GTGGAGAACGCTTTTCCGATGATGTTGGGCGCATGAAACACCTTCAGATTGACGCTC 567  
 61 ATAGACTCGGTTTGGACCTTGACGCTTCACTTCTTGGAACTGCTCCCAATATGAC 120  
 568 ATAGACTCGGTTTGGACCTTGACGCTTCACTTCTTGGAACTGCTCCCAATATGAC 627  
 121 AACCTCAATCAAGGCTATGACAGACCTTGAAGCGGCTCGAAGCAAGGCCAATGCGAT 180  
 628 AACCTCAATCAAGGCTATGACAGACCTTGAAGCGGCTCGAAGCAAGGCCAATGCGAT 687  
 181 GAATGGGTATTAATCTGAGAGAGGCACTGAGAGGCGATTCGGTAGCCGTTATTTGTCC 240  
 688 GAATGGGTATTAATCTGAGAGAGGCACTGAGAGGCGATTCGGTAGCCGTTATTTGTCC 747  
 241 GATTTTCTCTGCGGGGCTTCTTGGGCAAGGTCGCTGCGTGGCAATCAATGAAGGCG 300  
 748 GATTTTCTCTGCGGGGCTTCTTGGGCAAGGTCGCTGCGTGGCAATCAATGAAGGCG 807  
 301 ATTCAACCGGCGACAGAGCTGAAACTCCGACTGCTGCTCCCTGCTTCCGGTGGTGGC 360  
 808 ATTCAACCGGCGACAGAGCTGAAACTCCGACTGCTGCTCCCTGCTTCCGGTGGTGGC 867  
 361 CGCATGCAAGAGACATGAGCTTTGTGATGATGTTGTCATTAACCGGCGCTGTGCG 420  
 868 CGCATGCAAGAGACATGAGCTTTGTGATGATGTTGTCATTAACCGGCGCTGTGCG 927  
 421 CGTCAACCGGCGACAGAGCTTGGCGTCCGCTGCTGATGTTGCGGAAATCCCAATGGTGGC 480  
 928 CGTCAACCGGCGACAGAGCTTGGCGTCCGCTGCTGATGTTGCGGAAATCCCAATGGTGGC 987  
 481 GCCATGCGCTCGTGGGCTTCACTGAGGACATCTCACTTTTGGCGAAACCGGCGCGAGATA 540  
 988 GCCATGCGCTCGTGGGCTTCACTGAGGACATCTCACTTTTGGCGAAACCGGCGCGAGATA 1047  
 541 GGTTCCTGCTGCTGCGTGGTGAATTAACCACTGGGCAATGCGCTTTCAGACGCTGTG 600  
 1048 GGTTCCTGCTGCTGCGTGGTGAATTAACCACTGGGCAATGCGCTTTCAGACGCTGTG 1107  
 601 CAGCAGCGGCGAATTTGGTGAATACTGCTGATGATGAAATTTGTGCGCACTCCAA 660  
 1108 CAGCAGCGGCGAATTTGGTGAATACTGCTGATGATGAAATTTGTGCGCACTCCAA 1167  
 661 TTGCGTGAAGCGGTGGCAAAACCTCAAGTTATTCAGCGGCTGAGGCGAAGGATGCT 720  
 1168 TTGCGTGAAGCGGTGGCAAAACCTCAAGTTATTCAGCGGCTGAGGCGAAGGATGCT 1227  
 721 TTTTCTCCAACAACCTCTGCGGTGGCACTTCGGGTGATGAGCGAATTCGCGTTCTCGT 780  
 1228 TTTTCTCCAACAACCTCTGCGGTGGCACTTCGGGTGATGAGCGAATTCGCGTTCTCGT 1287  
 781 GACCGCGAGAGCGCTGGAATTCGGGGAATTAATGAAACCTTGGGGCGACAGCTGCTCAAG 840  
 1288 GACCGCGAGAGCGCTGGAATTCGGGGAATTAATGAAACCTTGGGGCGACAGCTGCTCAAG 1347  
 841 CTTTCTGGTGGCGCTGCGCATTTGAGCCCGGCTGTGCGGCTTGGCGCGCATC 900  
 1348 CTTTCTGGTGGCGCTGCGCATTTGAGCCCGGCTGTGCGGCTTGGCGCGCATC 1407  
 901 GGGGGCGCGCGCTGTGTGATGATGAGATGCGCGCTTTCAGCTTGGCGCGAGAG 960  
 1408 GGGGGCGCGCGCTGTGTGATGATGAGATGCGCGCTTTCAGCTTGGCGCGAGAG 1467

961 CTGCGTTTGGCGGTGCGGCGCATTTGCTGAGGCGCGAGCTAAACCTGCGGATCGGTCC 1020  
 1468 CTGCGTTTGGCGGTGCGGCGCATTTGCTGAGGCGCGAGCTAAACCTGCGGATCGGTCC 1527  
 1021 ATCATGACACCTCCGCGCGCATTTGTGAGGCGGCTGAGAGCTCGGATCGCAAC 1080  
 1528 ATCATGACACCTCCGCGCGCATTTGTGAGGCGGCTGAGAGAGCTCGGATCGCAAC 1587  
 1081 TCGATTGCGCGACCTTGTCCAGCTTATGACGCTCCCTCCCAACCGTTTGGTCAAT 1140  
 1588 TCGATTGCGCGACCTTGTCCAGCTTATGACGCTCCCTCCCAACCGTTTGGTCAAT 1647  
 1141 ATTGGTCAGGGGCTTGGGCGGTGGCGCGCATTTGCGGCGGATCGGATCGTCAAGG 1200  
 1648 ATTGGTCAGGGGCTTGGGCGGTGGCGCGCATTTGCGGCGGATCGGATCGTCAAGG 1707  
 1201 GCCGAAACGCGTGGCTGTCCGCAATTTGCAACAGAGGCGGCTCGGCAATCTCTTCCGC 1260  
 1708 GCCGAAACGCGTGGCTGTCCGCAATTTGCAACAGAGGCGGCTCGGCAATCTCTTCCGC 1767  
 1261 GACACCAACACGCGCGGAAATCAATAGCGACAGAGGCGCTGACGCGACGCACTTTTA 1320  
 1768 GACACCAACACGCGCGGAAATCAATAGCGACAGAGGCGCTGACGCGACGCACTTTTA 1827  
 1321 AGCCAAAGGCTTATGAGAGGATTCGTCGCCGAAACCGCAACTTTTGAATAATTTCTC 1380  
 1828 AGCCAAAGGCTTATGAGAGGATTCGTCGCCGAAACCGCAACTTTTGAATAATTTCTC 1887  
 1381 GGCACAAATCAGAACGCGCTCTCCGAATTTGATTAACAATCCGAGAGGCGGAGCGGAC 1440  
 1888 GGCACAAATCAGAACGCGCTCTCCGAATTTGATTAACAATCCGAGAGGCGGAGCGGAC 1947  
 1441 AGTCGCTTCAACGATTTGAAGGCTTTAGCGCAG 1473  
 1948 AGTCGCTTCAACGATTTGAAGGCTTTAGCGCAG 1980

RESULT 7  
 AR202319 2123 bp DNA linear PAT 20-APR-2002  
 LOCUS AR202319  
 DEFINITION Sequence 1 from patent US 6361986.  
 ACCESSION AR202319  
 VERSION AR202319.1 GI:20256858  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 2123)  
 AUTHORS Tlig,Y., Eikmanns,B., Eggeling,L., Sahm,H. and Mockel,B.  
 TITLE Process for the preparation of L-amino acids by fermentation and  
 nucleotide sequences coding for the acdA gene  
 JOURNAL Patent: US 6361986-A 1 26-MAR-2002;  
 FEATURES  
 source 1..2123  
 location/Qualifiers  
 BASE COUNT 460 a 574 c 592 g 497 t  
 ORIGIN  
 Query Match 100.0%; Score 1473; DB 6; Length 2123;  
 Best Local Similarity 100.0%; Pred. No. 6,4e-293;  
 Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTGAGAACGCTTTTCCGATGATGTTGGGCGCATGAAACACCTTCAGATTGACGCTC 60  
 508 GTGAGAACGCTTTTCCGATGATGTTGGGCGCATGAAACACCTTCAGATTGACGCTC 567  
 61 ATAGACTCGGTTTGGACCTTGACGCTTCACTTCTTGGAACTGCTCCCAATATGAC 120  
 568 ATAGACTCGGTTTGGACCTTGACGCTTCACTTCTTGGAACTGCTCCCAATATGAC 627  
 121 AACCTCAATCAAGGCTATGACAGACCTTGAAGCGGCTCGAAGCAAGGCCAATGCGAT 180

```

Db 628 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGGCTGAAAGCAAGGCCAATGCGAT 687
Qy 181 GAATGGTAAATTAATGAGAAAGGCAACCGTGGAGGGCATTCGCGTACCGCTTATTTGTC 240
Db 688 GAATGGTAAATTAATGAGAAAGGCAACCGTGGAGGGCATTCGCGTACCGCTTATTTGTC 747
Qy 241 GATTTTCTCTCTCGGCGGTTCTTTGGGCAAGGTCGCGTGGTGGCATCATGAAGGCG 300
Db 748 GATTTTCTCTCTCGGCGGTTCTTTGGGCAAGGTCGCGTGGTGGCATCATGAAGGCG 807
Qy 301 ATTCAACCGGCGCAAGAGCTGAACTCCACACTGCTGCTCCCTGCTTCCGCGTGGTGG 360
Db 808 ATTCAACCGGCGCAAGAGCTGAACTCCACACTGCTGCTCCCTGCTTCCGCGTGGTGG 867
Qy 361 GCGAATGAGGAGAAATGAGAGCTTTTGTGATGATGATGATGATGATGATGATGATGATGAT 420
Db 868 GCGAATGAGGAGAAATGAGAGCTTTTGTGATGATGATGATGATGATGATGATGATGATGAT 927
Qy 421 GGTCAACCGGCGCAAGAGCTGAACTCCACACTGCTGCTCCCTGCTTCCGCGTGGTGG 480
Db 928 GGTCAACCGGCGCAAGAGCTGAACTCCACACTGCTGCTCCCTGCTTCCGCGTGGTGG 987
Qy 481 GCGATGCGCTCGTGGGCGTTCATCTGGGCACTCTGCTTGGCGAAACCGGCGCGAGATA 540
Db 988 GCGATGCGCTCGTGGGCGTTCATCTGGGCACTCTGCTTGGCGAAACCGGCGCGAGATA 1047
Qy 541 GGTTCCTGCGTCCGCGTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Db 1048 GGTTCCTGCGTCCGCGTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1107
Qy 601 CAGCAGGCGGAGAAATTTGGTGAATACTGGTGTGATGATGATGATGATGATGATGATGATGAT 660
Db 1108 CAGCAGGCGGAGAAATTTGGTGAATACTGGTGTGATGATGATGATGATGATGATGATGATGAT 1167
Qy 661 TTGCGTGAAGCGGCGGCGGCAAAACCTCAAGGTTATTCAGCGGCGGAGCGCAACGATGCT 720
Db 1168 TTGCGTGAAGCGGCGGCGGCAAAACCTCAAGGTTATTCAGCGGCGGAGCGCAACGATGCT 1227
Qy 721 TTTTCTCAACAACCTCTGGGCGTGGCACTTCGCGTGTGATGAGAGGCAATTCGCGGCTCTG 780
Db 1228 TTTTCTCAACAACCTCTGGGCGTGGCACTTCGCGTGTGATGAGAGGCAATTCGCGGCTCTG 1287
Qy 781 GACCCGAGAGGCGCTGGAATCGGGGAGATTATGAGAAAGTTGGGGGAGAGAGCTGCTCAAG 840
Db 1288 GACCCGAGAGGCGCTGGAATCGGGGAGATTATGAGAAAGTTGGGGGAGAGAGCTGCTCAAG 1347
Qy 841 CTTTCTGCTGCGGCTGCTGGCGCATTTGAGCCGCGCTGTGCGGCTTCCCTGGCGCGCATC 900
Db 1348 CTTTCTGCTGCGGCTGCTGGCGCATTTGAGCCGCGCTGTGCGGCTTCCCTGGCGCGCATC 1407
Qy 901 GGGGGCGGGCGCGTGGTGTGATTTGGGAGAGATGCGCGCTTCAAGCTTGGGCGGAGAGAG 960
Db 1408 GGGGGCGGGCGCGTGGTGTGATTTGGGAGAGATGCGCGCTTCAAGCTTGGGCGGAGAGAG 1467
Qy 961 CTGCGTTTTTGGCGCTGTGGCATTTTGGCGCGCGAGCTTAAACCTGCGCATCTGTGTC 1020
Db 1468 CTGCGTTTTTGGCGCTGTGGCATTTTGGCGCGCGAGCTTAAACCTGCGCATCTGTGTC 1527
Qy 1021 ATCATGACACCTCGCGGCGCGGAGATTGTGCGAGGCGGCTGAGAGAGCTGGGATCGCAAGC 1080
Db 1528 ATCATGACACCTCGCGGCGCGGAGATTGTGCGAGGCGGCTGAGAGAGCTGGGATCGCAAGC 1587
Qy 1081 TCGATTGGCGGACCTGTGCAAGCTTATGCAAGCTTCCCGCCCAACCGTTCCGTTCAAT 1140
Db 1588 TCGATTGGCGGACCTGTGCAAGCTTATGCAAGCTTCCCGCCCAACCGTTCCGTTCAAT 1647
Qy 1141 ATTGGTCAAGGCGTGTGGCGGTGGCGCGTGGCATGCTGCGCGCATCTGTTCAACGCG 1200
Db 1648 ATTGGTCAAGGCGTGTGGCGGTGGCGCGTGGCATGCTGCGCGCATCTGTTCAACGCG 1707
Qy 1201 GCGGAAAACGCGTGGCTGTGCGCATTTGCGACACGAGGCGCGCTGGGCATCTCTTCCG 1260
Db 1708 GCGGAAAACGCGTGGCTGTGCGCATTTGCGACACGAGGCGCGCTGGGCATCTCTTCCG 1767

```

```

Qy 1261 GACACCAACCAAGCGCGGGAATTCATAGAGGCAAGGCGTGGAGGCGGCACTTTTA 1320
Db 1768 GACACCAACCAAGCGCGGGAATTCATAGAGGCAAGGCGTGGAGGCGGCACTTTTA 1827
Qy 1321 AGCCAAAGGCGTTATCGACGGGATGCTGCGGAAACCGAGACATTTGTTGAAGAAATTC 1380
Db 1828 AGCCAAAGGCGTTATCGACGGGATGCTGCGGAAACCGAGACATTTGTTGAAGAAATTC 1887
Qy 1381 GGCACATCAAGCAAGCGCTCTCCGAATTGGATTAACATCCGAGAGGCGGAGCGGAC 1440
Db 1888 GGCACATCAAGCAAGCGCTCTCCGAATTGGATTAACATCCGAGAGGCGGAGCGGAC 1947
Qy 1441 AGTGGCTTCAACGATTTGAGCGGTTAGCGGAG 1473
Db 1948 AGTGGCTTCAACGATTTGAGCGGTTAGCGGAG 1980

RESULT 8
BD004741
LOCUS 2123 bp DNA linear PAT 31-JAN-2002
DEFINITION Replicable DNA, amino acid sequence, Corynebacterium microorganism,
shuttle vector, and process for producing L-amino acid.
ACCESSION BD004741
VERSION BD004741.1 GI:18632702
KEYWORDS JP 2001008693-A/1
SOURCE Corynebacterium glutamicum
ORGANISM Corynebacterium glutamicum
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1 (bases 1 to 2123)
AUTHORS Tilke, I., Eggering, L., Eickmans, B., Zamu, H. and Meck, V.
TITLE Replicable DNA, amino acid sequence, Corynebacterium microorganism,
shuttle vector, and process for producing L-amino acid
JOURNAL Patent: JP 2001008693-A 1 16-JAN-2001;
DECUSSA HUELS AG, FORSCHUNGSZENTRUM JUERDICH GMBH
COMMENT OS Corynebacterium glutamicum
PN JP 2001008693-A/1
PD 16-JAN-2001
PF 24-MAY-2000 JP 2000153547
PR 27-MAY-1999 DE 19924365.4
PI IVENNE TILKE, IOTHAR EGGERING, BERNHARD EICKMANS, HERMANN ZAMU,
PI VETRYNA MECKEL
PC C12N15/09, C12N1/21, C12P13/04, C12P13/06, C12P13/08, PC
C12P13/08
PC C12P13/12, C12P13/20, C12P21/02// (C12N15/09, C12R1:15), (C12N1/21,
PC C12R1:15),
PC C12P13/04, C12R1:15), (C12P13/06, C12R1:15), (C12P13/08, C12R1:15), PC
(C12P13/12, C12R1:15), (C12P13/20, C12R1:15), (C12P21/02, C12R1:15), PC
C12N15/00, C12R1:15)
PC (C12N15/00, C12R1:15)
CC CC
FH Key location/Qualifiers
FT gene (508).. (1980).
FEATURES
source 1..2123
/organism="Corynebacterium glutamicum"
/mol_type="genomic DNA"
/db_xref="taxon:1718"
BASE COUNT 460 a 574 c 592 g 497 t
ORIGIN
Query Match 100.0%; Score 1473; DB 6; Length 2123;
Best Local Similarity 100.0%; Pred. No. 6,4e-293;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GTGAGAGAGGTTTCCGACTATGATGTGGGAGATGGAACACACTCAGCATTGAGGCTC 60
Db 508 GTGAGAGAGGTTTCCGACTATGATGTGGGAGATGGAACACACTCAGCATTGAGGCTC 567
Qy 61 ATAGACTCGGTTTGAACCTCGACAGCTTCATTTCTTGAATGAATCCCAATATGAC 120

```

[illegible]

Oy		1201	GCCGAAAACGGTGGCTGTCCGCATTTCACACAGAGGCGCTCGGCATCTCTTCGCG	1260
Db		1708	GCCGAAAACGGTGGCTGTCCGCATTTCACACAGAGGCGCTCGGCATCTCTTCGCG	1767
Oy		1261	GACACCAACGCGCGGGAATCATAGCGACAGGCGTGACGCGCACACTTTTA	1320
Db		1768	GACACCAACGCGCGGGAATCATAGCGACAGGCGTGACGCGCACACTTTTA	1827
Oy		1321	AGCCAAGGCTTATTCGACGGGATTCGTGCCGAAACGACGACTTTGTGAAGAATCTC	1380
Db		1828	AGCCAAGGCTTATTCGACGGGATTCGTGCCGAAACGACGACTTTGTGAAGAATCTC	1887
Oy		1381	GGCACAATTCAGCAACGCCCTCTCCGAAATTGGATAATCCGAGAGGCGCGACGCGAC	1440
Db		1888	GGCACAATTCAGCAACGCCCTCTCCGAAATTGGATAATCCGAGAGGCGCGACGCGAC	1947
Oy		1441	AGTCGCTTCAACGAGATTGAGCGTTTAGCGCAG	1473
Db		1948	AGTCGCTTCAACGAGATTGAGCGTTTAGCGCAG	1980
RESULT 9				
AP005276/c		332050 bp	DNA	linear BCT 08-AUG-2002
LOCUS				
DEFINITION		Corynebacterium glutamicum ATCC 13032 DNA, complete genome, section		
VERSION		3/10.		
KEYWORDS		AP005276 BA000036		
SOURCE		AP005276.1 GI:21323419		
ORGANISM		Corynebacterium glutamicum ATCC 13032		
REFERENCE		Corynebacterium glutamicum ATCC 13032		
AUTHORS		Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
TITLE		Corynebacteriaceae; Corynebacterium.		
JOURNAL		Nakagawa,S. Complete genomic sequence of Corynebacterium glutamicum ATCC 13032 .Unpublished 2 (bases 1 to 332050)		
COMMENT		Direct Submission Submitted (24-MAY-2002) Satoshi Nakagawa, Kyowa Hakko Kogyo Co. Ltd., Tokyo Research Laboratories; 3-6-6, Asahi-machi, Machida, Tokyo 194-8533, Japan (E-mail:snakagawa@xanagen.com, Tel:81-44-829-3031, Fax:81-44-813-1651) This sequence is conducted by collaboration of Kyowa Hakko Kogyo Co. Ltd. And Kitaseito University. Location/Qualifiers 1. .332050 /organism="Corynebacterium glutamicum ATCC 13032" /mol_type="genomic DNA" /strain="ATCC 13032" /db_xref="taxon:1196627" /note="ATCC 13032" 314. .2518 /gene="CG10654" 314. .2518 /gene="CG10654" /note="PF02554:Carbon starvation protein CstA" /codon_start=1 /transl_table=11 /product="Carbon starvation protein, predicted membrane protein" /protein_id="PBAB98047.1" /db_xref="GI:21323420" /translation="MRKSPMSPTARVGLLVFGVIAVGWGAIAFSRGETTINSVMVL AAGSYIIAFSFARLIEYKVMXPKQRPATPYAVNDKDYPIDSRVILGFHFAIA GAPLPGVMAQMGLPGLIITLLIVAGADYLIVMSTRGRSLSGOVNRDM RGTVAAGLIATISIMIIIIVALLIVNVALDSDPWGSITWTPIIALPMGYLRLL RPRREVSIGTIVALLIIVGVGWADDSKWGVEMFTSTTALIGIMAAIIP VILLAPRYLSIFMKITGVYIGLAIVLTIRARPVNQPSVTSPLENGNPFSGLPEF LFITICAGLISGHFHALISSGTYTPKLVEKSGOMMLDYGWLMEFVAMMALITAVILD RHLYEFMNPLALTGDPTPAEMWNVISIGTMDIPDEQLSEBAESVSGESTIVISRGG		
FEATURES				
source				
gene				
CDS				

gene  
CDS  
2522. .2725  
/gene="Cg10655"  
/gene="Cg10655"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="BAB98048.1"  
/db\_xref="GI:21323421"  
/translacion="MMALFTALMKIPRAVWWYLTLMGDTAYSKYVYHLKHHNDAPI  
PTERYRYAKYDQDANGARC"  
complement(3092. .3703)  
/gene="Cg10656"  
complement(3092. .3703)  
/gene="Cg10656"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="BAB98049.1"  
/db\_xref="GI:21323422"  
/translacion="MSNVIINLVLMQASHTHOVNLIIPSSDATDQIILATSSILCSQ  
NRRNIISWGVPRSEERVWEIKLTHRSVSVKSPFRGITSQTNCRCTPIGNMDDIC  
HRTGANSSTHSGSCILHRCIVDDAVDHFILSVRSILNMVGNASDILVSQVLLVELFER  
RTVGAFFVNVHGMVTPSQRTNVVAHTNSLSKSNVGHCKSCVIL"  
3750. .4706  
/gene="Cg10657"  
3750. .4706  
/gene="Cg10657"  
/codon\_start=1  
/transl\_table=11  
/product="Unclassified protein involved in propionate  
catabolism"  
/protein\_id="BAB98050.1"  
/db\_xref="GI:21323423"  
/translacion="MMNTIYQALIGQALHTTTATROSKAISWKAFAFAPAGMSIE  
AYDRAMRGASPIWBGEDGVIAWLISGLDHYITIPLPARGAKRALIDVYTHESHA  
EYQSAQAPIDLASMGKELAAQSLDRDVSIVLHSHHTHYIVIGTSGNDPKPPDMS  
RETLHSHIVITFAVLEDRAMHERSVAPESRAHRETIELMNKISTVEDPWTHTV  
VDPAKAFARAVITFEKGTVEDELAVAADHPICARPAEYOYIOKRTPLAEVSEEB  
KEODRFLDAQRTHLEDLSEINIELDADIILAKAPIPEGLF"  
4715. .5629  
/gene="Cg10658"  
4715. .5629  
/gene="Cg10658"  
/codon\_start=1  
/transl\_table=11  
/product="PEP phosphonmutase and related enzymes"  
/protein\_id="BAB98051.1"  
/db\_xref="GI:21323424"  
/translacion="MFSSAVAPTEKRALRALAAPETARMGAFSPILAARAOEAG  
BEVYSGAVAADALPIDGLITTEVAHRSQIARVDLVLVDADFGCEPMSAAR  
TUSLEDJAGACHLIEDOVNPKRCGLHGEKEVGVISIAAVNERSEOCVICA  
RTDAGVESIDSAIERAKAYADAGADMITEALVSPADFEKFRAAVIDPILANNTEG  
KTELPACOLLEDIGYNAVITYPTLLRIMGQVEQALSDIANTGICTDVMDMQHSRL  
YELLRINEINADQOFTYSADSYKPIF"  
5658. .6809  
/gene="Cg10659"  
5658. .6809  
/gene="Cg10659"  
/EC\_number="4.1.3.7"  
/note="PFO0285:Citrate synthase"  
/codon\_start=1  
/transl\_table=11  
/product="Citrate synthase"  
/protein\_id="BAB98052.1"  
/db\_xref="GI:21323425"  
/translacion="MSSATTTDVVRKGLYGVIAIDTAVSKWPEPNSLTYRGVAVEDLV  
ENCSEEEVYVLLMHBELPTAQQLAEFNRGSGSYSLDAGLISLHSLPKEHMPDVMWR

gene  
CDS      complement (6846 . 7040)  
         /gene="Cg10660"  
         /codon\_start=1  
         /transl\_table=1  
         /product="Hypothetical protein"  
         /protein\_id="BAB98053\_1"  
         /db\_xref="GI:21323428"  
         /translation="MYFDVSDQVDDELHFSEFFELYADQAALDAHRCAPHFYVMREAVA  
KHLVPGSKXNVGTFRMLHHG"  
         /complement (7129 . 8172)  
         /gene="Cg10661"  
         complement (7129 . 8172)  
         /gene="Cg10661"  
         /EC\_number="1.1.1.82"  
         /note="PF02615:Malate/L-lactate dehydrogenase"  
         /codon\_start=1  
         /transl\_table=1  
         /product="Malate/L-lactate dehydrogenases"  
         /protein\_id="BAB98054\_1"  
         /db\_xref="GI:21323427"  
         /translation="MPPEVTVAQQLTFTLCTDILTKTGPAPADAHLVGDSIVQADLMGH  
PSHGSLRLPWVRRLSHGVMTATHAVLVNDLGVALMDGNIGIQLVLADHARKEAYATW  
RAMMGAGVASVRNSNHGTAMYTTRAAAGCVSILITNNAAPMAAPMGGSEKIINN  
PMSIAAPEGAFETRAVDIANTRAVARCKTVHAKOTMWPIDETUALTSBAATPPEDPREAIN  
GVNLPMAGHKRYAISFMWDVLSGVLTSSQSHTKYNGEYDPPTPGGAHFLTADLYAAP  
RDPDFDALSDILEVKSTPKAQNTBEIFYPGESBDRAHRKNASHGISPEKTMWEL  
QELAIEHVNVTHR"

gene  
CDS      complement (8265 . 8987)  
         /gene="Cg10662"  
         complement (8265 . 8987)  
         /note="PF00392:Bacterial regulatory proteins, gntR family"  
         /codon\_start=1  
         /transl\_table=1  
         /product="Transcriptional regulators"  
         /protein\_id="BAB98055\_1"  
         /db\_xref="GI:21323428"  
         /translation="MVQCDAQSPATRKAQOVYTIQRRIEDGTLPNGRMSEYMVLVE  
HTGSRTVPRALRRLLADELIILEPROAPVSPISRHIKIDLEFPRIYEVALBETI  
SVAGSGRIRGEFSTLAADRELLENAGADAFPADRRLTKSDPLVAANTHQIFEG  
RSLSIKRHTTRLRIIASHPBARLRQSVQHIEENCEAVASGDLSRNAACROHLIHVB  
KSIIIALINADSTSGCIDTRS"

gene  
CDS      complement (9094 . 9798)  
         /gene="Cg10663"  
         complement (9094 . 9798)  
         /gene="Cg10663"  
         complement (9094 . 9798)  
         /codon\_start=1  
         /transl\_table=1  
         /product="Hypotheical membrane protein"  
         /protein\_id="BAB98056\_1"  
         /db\_xref="GI:21323429"  
         /translation="MNPSSRVALLTTAVSIPTASAATLTADTDKELCIANSDDSA  
VVTNRSEDVSERQRLDELDPAODPGIKAIESTIADDNAPTAELIQYLAESESE  
GLIAMLLPDPPDLADPNABESEFKTEITYDEAKDIISGSSPASVLSQLOOAPTGTGR  
TABRIAREVFARTDYNESQTALKEPONCIDAIDARPPIPLYILLIGAWATAVIYAL  
GIRAWINSRKOSKHSQ"

gene  
CDS      complement (9878 . 12094)  
         /gene="Cg10664"  
         complement (9878 . 12094)  
         /gene="Cg10664"  
         /EC\_number="1.1.1.42"  
         /note="FIG00178: monomer \_idh: isocitrate dehydrogenase,  
NADP-dependent"  
         /codon\_start=1  
         /transl\_table=1  
         /product="Monomeric isocitrate dehydrogenase (NADP+)"



/protein\_id="BAB98057.1"  
 /db\_xref="GI:21323430"  
 /translation="MAKIMTRDDEAPLATVSLKRVENFAPATAGIEVETDIDSLAG  
 RIAOPERLTDOKVGNALIELGELAKTPEANITGLPISASPOLKAIKIELDQOG  
 YVILPELPNATDEEDDILARIYAVGSAVNPLRGNSDRRPIAVKPFKPFPHRM

Query Match 100.0%; Score 1473; DB 1; Length 332050;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-293;  
 Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTGAGAGAACGTTTCCGACTATGCTGTGGGCGATGAAACACCTTCAGATTGACGCTC 60
DB 213164 GTGAGAGAACGTTTCCGACTATGCTGTGGGCGATGAAACACCTTCAGATTGACGCTC 213105
QY 61 ATAGACTCGGTTTGGACCCCTGACAGCTTCATTTCTTGAATGAACTCCCAATATGAC 120
DB 213104 ATAGACTCGGTTTGGACCCCTGACAGCTTCATTTCTTGAATGAACTCCCAATATGAC 213045
QY 121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGCCAAATGCGAT 180
DB 213044 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGCCAAATGCGAT 212985
QY 181 GAATGGTATTAATCTGAGAGAGCGACCTGAGAGGCGATTCGGATGCGCTTATTTGTCC 240
DB 212984 GAATGGTATTAATCTGAGAGAGCGACCTGAGAGGCGATTCGGATGCGCTTATTTGTCC 212925
QY 241 GATTTTCTCTCGGCGGCTTCTTGGGCGAGCTGCGTGGTGGCATCATGAAAGCG 300
DB 212924 GATTTTCTCTCGGCGGCTTCTTGGGCGAGCTGCGTGGTGGCATCATGAAAGCG 212865
QY 301 ATTCAACCGGCGACAGAGCTGAAACTCCGACTGCTGCTCCCTGCTTCCGCTGGTGG 360
DB 212864 ATTCAACCGGCGACAGAGCTGAAACTCCGACTGCTGCTCCCTGCTTCCGCTGGTGG 212805
QY 361 CGCATGCGAGAGACATGAGACTTTTGTATGATGCTGTCCATTAACCGGCGCTGTGCG 420
DB 212804 CGCATGCGAGAGACATGAGACTTTTGTATGATGCTGTCCATTAACCGGCGCTGTGCG 212745
QY 421 CGTACCGGAGAGGCGATTTGCGGTTCTCGGTGATTTGGCGAATCCCATGAGTGGTGG 480
DB 212744 CGTACCGGAGAGGCGATTTGCGGTTCTCGGTGATTTGGCGAATCCCATGAGTGGTGG 212685
QY 481 GCCATGCGCTGTGGGTTCACTGAGGACCTCACTTTTCCGAAACCGGCGCGCAGATA 540
DB 212684 GCCATGCGCTGTGGGTTCACTGAGGACCTCACTTTTCCGAAACCGGCGCGCAGATA 212625
QY 541 GGTTCCTGCGGTTCTCGGCTGTGATTAACCACTGGGCGATGCCCTTCAGACGCTGTG 600
DB 212624 GGTTCCTGCGGTTCTCGGCTGTGATTAACCACTGGGCGATGCCCTTCAGACGCTGTG 212565
QY 601 CAGCAGGCGAGAAATTTGTTGAAAATGTTGATTTGAAATTTGTTGCTGCTCACTCCA 660
DB 212564 CAGCAGGCGAGAAATTTGTTGAAAATGTTGATTTGAAATTTGTTGCTGCTCACTCCA 212505
QY 661 TTGCGTGCAGCGGTGGCAAAAACCTCAAGTTATTCAGCCGCTGAGGCAACGATGCT 720
DB 212504 TTGCGTGCAGCGGTGGCAAAAACCTCAAGTTATTCAGCCGCTGAGGCAACGATGCT 212445
QY 721 TTTTCTCCAACTCTCGGCGTGGCACTTCGCGTGTATGAGCGAATTCGCGTCTCGT 780
DB 212444 TTTTCTCCAACTCTCGGCGTGGCACTTCGCGTGTATGAGCGAATTCGCGTCTCGT 212385
QY 781 GACCGCGAGAGCGCTGGAATCGGGAGATTAATGAAAAGTTGGGGCGAGAGTCTCAAG 840
DB 212384 GACCGCGAGAGCGCTGGAATCGGGAGATTAATGAAAAGTTGGGGCGAGAGTCTCAAG 212325
QY 841 CTTTCTGTGCGCTGTGCGCATTTGAGCCCGCTGTGCGGCTTCCGCGCGCATC 900
DB 212324 CTTTCTGTGCGCTGTGCGCATTTGAGCCCGCTGTGCGGCTTCCGCGCGCATC 212265
QY 901 GGGGGCGGCGCGTGTGCTGATTTGGGAGATTCGCGCTTCAACGCTTGGGCGGAGAG 960
DB 212264 GGGGGCGGCGCGTGTGCTGATTTGGGAGATTCGCGCTTCAACGCTTGGGCGGAGAG 212205

```

```

QY 961 CTGCGTTTGGCGGTGTGGCAATTTGCTGGCGCGGAGCTAAACCTGCGATGCTGTC 1020
DB 212204 CTGCGTTTGGCGGTGTGGCAATTTGCTGGCGCGGAGCTAAACCTGCGATGCTGTC 212145
QY 1021 ATCATGACACCTCCGCGCGCAATTTGTGACAGCGGCTGAGAGCTCGGCATCGAAGC 1080
DB 212144 ATCATGACACCTCCGCGCGCAATTTGTGACAGCGGCTGAGAGCTCGGCATCGAAGC 212085
QY 1081 TCGATTGCGCGCACTTGTCCAAAGCTTATGACGCTCCCTCCCAACCGTTTGGTCAAT 1140
DB 212084 TCGATTGCGCGCACTTGTCCAAAGCTTATGACGCTCCCTCCCAACCGTTTGGTCAAT 212025
QY 1141 ATTGTGACAGGCGTTTGGCGGTGGCGGCTGCGGCTGCGCATCTCCGCGCATCTGCTACGG 1200
DB 212024 ATTGTGACAGGCGTTTGGCGGTGGCGGCTGCGGCTGCGCATCTCCGCGCATCTGCTACGG 211965
QY 1201 GCCGAAAACGCGTGTGCTGTCCGCAATTTGACAGAGAGGCGCTGGGCGATCTTCGCG 1260
DB 211964 GCCGAAAACGCGTGTGCTGTCCGCAATTTGACAGAGAGGCGCTTCGCGCATCTTCGCG 211905
QY 1261 GACACCAACCAAGCGCGCGGAAATCATAGAGCAAGGCGTGCAGGCGCAGCATTTTA 1320
DB 211904 GACACCAACCAAGCGCGCGGAAATCATAGAGCAAGGCGTGCAGGCGCAGCATTTTA 211845
QY 1321 AGCCAAAGGCTTATGAGAGGAGTCTGCGCCGCAACCGAGCACTTTGTTGAATAATCTC 1380
DB 211844 AGCCAAAGGCTTATGAGAGGAGTCTGCGCCGCAACCGAGCACTTTGTTGAATAATCTC 211785
QY 1381 GGCACCAATCAGCAAGCCCTCTCCGAATTTGATTAACAATCCGAGAGGCGGGAAGCGCAG 1440
DB 211784 GGCACCAATCAGCAAGCCCTCTCCGAATTTGATTAACAATCCGAGAGGCGGGAAGCGCAG 211725
QY 1441 AGTCGCTTCAACGATTTGAGGCTTTAGCGCAG 1473
DB 211724 AGTCGCTTCAACGATTTGAGGCTTTAGCGCAG 211692

```

```

RESULT 10
AX127145/c 349980 bp DNA linear PAT 11-MAY-2001
LOCUS
DEFINITION
Sequence 7061 from Patent EP1108790.
ACCESSION
AX127145 AX14121
VERSION
AX127145.1 GI:14041133
KEYWORDS
Corynebacterium glutamicum
SOURCE
Corynebacterium glutamicum
ORGANISM
Corynebacterium glutamicum
Bacteria; Actinobacteriia; Actinomycetales;
Corynebacteriaceae; Corynebacteriales; Corynebacterium.
REFERENCE
1 Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
Novel polynucleotides
Patent: EP 1108790-A 7061 20-JUN-2001;
JOURNAL
KYOMA HAKKO KOGYO CO., LTD. (JP)
FEATURES
source
location/Qualifiers
1..349980
/organism="Corynebacterium glutamicum"
/mol_type="genomic DNA"
/db_xref="taxon:1718"
/notice="Seq 1 to long (3,309,400) split in 11, seq 7061
0.600.001 0.949.980 ."
BASE COUNT
79274 a 90638 c 98727 g 81341 t
ORIGIN

```

Query Match 100.0%; Score 1473; DB 6; Length 349980;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-293;  
 Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTGAGAGAGCGTTTCCGACTATGCTGTGGGCGATGAAACACCTTCAGATTGACGCTC 60
DB 281114 GTGAGAGAGCGTTTCCGACTATGCTGTGGGCGATGAAACACCTTCAGATTGACGCTC 281055

```



QY	61	TTAGACTGGCTTTTGGACCTTGACAGCTCATTTTCTTTGGAGAAGAACTCCCAATATGAC	120
Db	281054	ATTGACTGGCTTTTGGACCTTGACAGCTCATTTTCTTTGGAGAAGAACTCCCAATATGAC	280995
QY	121	AACCTCAATCAAGGCGTATGACAGACCTTGGAGCGGGCTGAAAGCAAGGCCAATACGAT	180
Db	280994	AACCTCAATCAAGGCGTATGACAGACCTTGGAGCGGGCTGAAAGCAAGGCCAATACGAT	280935
QY	181	GAATGGTAAATTACTGGAGAAGGCAACGCTGGAGGGCAATTCGGTAGCCGTAATTTTGTCC	240
Db	280934	GAATGGTAAATTACTGGAGAAGGCAACGCTGGAGGGCAATTCGGTAGCCGTAATTTTGTCC	280875
QY	241	GATTTTTCCTTCCTCGGGCGGCTCTTTGGGACGCTGCGCTGGGCGATCATGAAGAGGCG	300
Db	280874	GATTTTTCCTTCCTCGGGCGGCTCTTTGGGACGCTGCGCTGGGCGATCATGAAGAGGCG	280815
QY	301	ATTACCGGCGCACAGAGCTGAACCTCCCACTGCTGCTCCCTCGCTTCCTCGGTGGGCG	360
Db	280814	ATTACCGGCGCACAGAGCTGAACCTCCCACTGCTGCTCCCTCGCTTCCTCGGTGGGCG	280755
QY	361	CGCATGCAAGAAACAATCGAGCTTTTGTGATGATGCTGTCCATAACCGCGGCGTGTCCAG	420
Db	280754	CGCATGCAAGAAACAATCGAGCTTTTGTGATGATGCTGTCCATAACCGCGGCGTGTCCAG	280695
QY	421	CGTCAACCGGAGAGCGCATTTTGCCTTCCTGGTGTATTTTGGCGAATCCCAAGATGGGTGC	480
Db	280694	CGTCAACCGGAGAGCGCATTTTGCCTTCCTGGTGTATTTTGGCGAATCCCAAGATGGGTGC	280635
QY	481	GCCATAGGCGCTCGTGGGGGTTCACTCTGGGGCAATCACTTTTGGCGAACCGCGCGCGAGATA	540
Db	280634	GCCATAGGCGCTCGTGGGGGTTCACTCTGGGGCAATCACTTTTGGCGAACCGCGCGCGAGATA	280575
QY	541	GGTATTCCTGGGTCCTCGCGGTGTGAATTAAACCATGCGGAGTACGCTTCACAGCGGTGTG	600
Db	280574	GGTATTCCTGGGTCCTCGCGGTGTGAATTAAACCATGCGGAGTACGCTTCACAGCGGTGTG	280515
QY	601	CAGCAGGCGGGAATTTTGTGAAAACTGGTGTGATTTGATGAGAATTTGTGTGCCACTTCCA	660
Db	280514	CAGCAGGCGGGAATTTTGTGAAAACTGGTGTGATTTGATGAGAATTTGTGTGCCACTTCCA	280455
QY	661	TTGCGGTCAGCGGTGGCAAAAAACCTCAAGGTTAATCAGCCGGTAAAGGCAACGATTCGT	720
Db	280454	TTGCGGTCAGCGGTGGCAAAAAACCTCAAGGTTAATCAGCCGGTAAAGGCAACGATTCGT	280395
QY	721	TTTTCTCCAACAACCTCTGGCGTGGCACTTCGCGTATGAGGCGAATTGCGCTTCTCGT	780
Db	280394	TTTTCTCCAACAACCTCTGGCGTGGCACTTCGCGTATGAGGCGAATTGCGCGTCTCGT	280335
QY	781	GACCCGCAAGAGGCTTGGAAATCGGGAGATTATGAAACCTTTGGGGGCAAGCTGTCAAG	840
Db	280334	GACCCGCAAGAGGCTTGGAAATCGGGAGATTATGAAACCTTTGGGGGCAAGCTGTCAAG	280275
QY	841	CTTTCGAGGCGCGTCTGGCGATTAAGGCGCGGCTGTGGCGGCTTCCCTCGGGGCGCATC	900
Db	280274	CTTTCGAGGCGCGTCTGGCGATTAAGGCGCGGCTGTGGCGGCTTCCCTCGGGGCGCATC	280215
QY	901	GGGGGCGGCGCGGTGTGCTGATTTGGGACAGAAATCGCGCTTCAACGCTTTGGGCGCAGAG	960
Db	280214	GGGGGCGGCGCGGTGTGCTGATTTGGGACAGAAATCGCGCTTCAACGCTTTGGGCGCAGAG	280155
QY	961	CTGCGTTTGGCGGTGTGAGCAATTTGCTGTGCGCGCGAGCTAAACCTGCCGATCGTGTCC	1020
Db	280154	CTGCGTTTGGCGGTGTGAGCAATTTGCTGTGCGCGCGAGCTAAACCTGCCGATCGTGTCC	280095
QY	1021	ATCATTCGAAACCTTCGGGCGGCCAAATTTGTGTGCAAGCGGCTGAAGAGCTCGGCATTCGAAGC	1080
Db	280094	ATCATTCGAAACCTTCGGGCGGCCAAATTTGTGTGCAAGCGGCTGAAGAGCTCGGCATTCGAAGC	280035
QY	1081	TGATTTGGCGCGCACTTTGTCCAAGCTTATTCAGAGCTCCCTCCCAACCGTTTGGGTCAATT	1140
Db	280034	TGATTTGGCGCGCACTTTGTCCAAGCTTATTCAGAGCTCCCTCCCAACCGTTTGGGTCAATT	279975
QY	1141	ATTGTTCAGAGGCGGTGGCGGTGGCGGCTGGCAATCTGTCCCGCGCATCTGATCTACGCG	1200

Db	279974	ATTGCTCAGGCGCTGGCGCGTGGCGCGCTGGCGCATGCTGCCCGCATCTGGCTTACGGG	279915
Qy	1201	GCCTGAAAACGCGTGGCTGTCCGCATTTGCCACAGAGGGCGCCTCGGCAATCTCTTCCGC	1260
Db	279914	GCCGAAAACGCGTGGCTGTCCGCATTTGCCACAGAGGGCGCCTCGGCAATCTCTTCCGC	279855
Qy	1261	GACACCAACACGCCCGCGGAAATCATTAAGCGACAAAGCGTTCAGAGCGGACCGACCTTTA	1320
Db	279854	GACACCAACACGCCCGCGGAAATCATTAAGCGACAAAGCGTTCAGAGCGGACCGACCTTTA	279795
Qy	1321	AGCCAAGGGCTTATCGACGGGATCGTCGCCGAAACCGAGCATTTGTGTAAGAAATTCTC	1380
Db	279794	AGCCAAGGGCTTATCGACGGGATCGTCGCCGAAACCGAGCATTTGTGTAAGAAATTCTC	279735
Qy	1381	GGCACAATCAGCAGCGCCCTCTCCGAATTGATTAACAATCCGAGAGGCGGGACGCGAC	1440
Db	279734	GGCACAATCAGCAGCGCCCTCTCCGAATTGATTAACAATCCGAGAGGCGGGACGCGAC	279675
Qy	1441	AGTCGCTTCACACGATTTGAGCGTTAAGGCGAG	1473
Db	279674	AGTCGCTTCACACGATTTGAGCGTTAAGGCGAG	279642

RESULT 11	AP005217/c	300750 bp	DNA	linear	BCT 20-DEC-2002
LOCUS	AP005217	300750 bp	DNA	linear	BCT 20-DEC-2002
DEFINITION	Corynebacterium efficiens YS-314 DNA, complete genome, section 4/11.				
ACCESSION	AP005217	BA000035			
VERSION	AP005217.1	GI:23492722			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE	1				
AUTHORS	Kawarabayashi, Y., Yamazaki, J., Hino, Y., Kikuchi, H., Nakamura, Y., Ikeo, K., Suzuki, M., Mashima, U., Itoh, T., Yamagishi, A., Nishio, Y., Usuda, Y. and Sugimoto, S.				
TITLE	The entire genomic sequence of Corynebacterium efficiens YS-314				
JOURNAL	Published Only in Database (2002)				
REFERENCE	2 (bases 1 to 300750)				
AUTHORS	Kawarabayashi, Y., Yamazaki, J., Hino, Y., Kikuchi, H. and Director-General of Biotechnology Center.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-MAY-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; Nishihara 2-49-10, Shibuya-ku, Tokyo 151-0066, Japan (E-mail: biotech@ite.go.jp, Tel:81-3-3461-1933, Fax:81-3-3461-6424)				
COMMENT	Kawarabayashi, Y. is officially affiliated with the National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, 305-8566 Japan				
	Nakamura, Y., Ikeo, K., Suzuki, M. and Mashima, J. are at the National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan				
	Itoh, T. is at the Japan Biological Information Research Center, Koto-ku, Tokyo, 135-0064 Japan				
	Yamagishi, A. is at Tokyo University of Pharmacy and Life Science, Hachioji, Tokyo, 192-0392 Japan				
	Nishio, Y., Usuda, Y. and Sugimoto, S. are at the Ajinomoto Co., Inc., Kawasaki, Kanagawa, 210-8691 Japan				
	The other authors are at the National Institute of Technology and Evaluation, Shibuya-ku, Tokyo, 151-0066 Japan.				
FEATURES	location/Qualifiers				
source	1..300750				
	/organism="Corynebacterium efficiens YS-314"				
	/mol_type="genomic DNA"				
	/strain="YS-314"				
	/db_xref="taxon:196164"				
	252_620				
	/note="CE0885"				
	/codon_start=1				
	/transl_table=11				
CDS					

[illegible]

/product="putative cold shock protein"  
 /protein\_id="BAC17705.1"  
 /db\_xref="GI:23492733"  
 /translation="MPVSTVKKYDPERGFVSNBGGDCVFGKQVLPKQVTEIHQGO  
 RIEFDPAQRGPOLARVLALETPEPRRPHOHTYPEELINGIISDMVTMLEGTVPALLK  
 GHYDPRKVGEOVAKILRAVAKLEBA"  
 complement (13881..14408)

Query Match 40.9%; Score 602.4; DB 1; Length 300750;  
 Best Local Similarity 63.9%; Pred. No. 5.8e-114;  
 Matches 928; Conservative 0; Mismatches 521; Indels 3; Gaps 1;

21 TATGTTGGGGCATGAGAACACTTTCAGCATGAGCTCATGAGCTGTTTGGACC 80  
 Db TAGTTTGAACCAAGACGAACTCCGCCAGGATCTGATCGATGATGATGAGCA 25999  
 26058  
 81 TGACAGCTTCATTCTTGGATGAACTCCCAATATGACAACTCAATCAAGCTATGC 140  
 Db GGACAGTTTCTGCTTCTGGGACACTCCCGGAGTACGGGGAGATGATGACTACCG 25939  
 25998  
 141 AGAGACTTGGAGCGGGCTCCGAGAGAGCCAAATCCGATGATGATGATGATGATG 200  
 Db GAGAGCGCTTGGCCCGCCGAGAGAAATCCGGGGTGAATGAGGATGATGATGATG 25879  
 25938  
 201 AGGCACTGGAGAGGCAATTCGGTACCGTATTTTTCGATTTTTCCTTCTCGGCG 260  
 Db GGGAACTGGAGAGGCTCCCGGTGGCTTTCATGCTTCCAGTTTCGCTTCTCGGCG 25819  
 25878  
 261 TTTCTTGGGCAAGTCCGCTGGTGGTGGATCATGAAAGGATTCACCGGCGACAGAGCT 320  
 Db TTGATTCGGTGGCGGACCTCCCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 25818  
 25818  
 331 GAAATCCCACTGCTGCTCTCCCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 380  
 Db AGCTTCCCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 25758  
 25758  
 381 AGCTTTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 440  
 Db GCGCTTCCCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 25698  
 25698  
 441 GCGCTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 500  
 Db GCGCTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 25638  
 25638  
 501 ATCTGGGCACTCACTTTTGGGAAACCGGCGCGAGATAGTTTCTTGGTCTCGGCT 560  
 Db GCGCGGCACTTCACTTATGCGGACCGGATGATGATGATGATGATGATGATGATG 25578  
 25578  
 561 GGTGAGTTTAACTGAGGATGCGCTTCAGACGGTGGTGGTGGTGGTGGTGGTGGTGG 620  
 Db GGTGAGTGGACCAACCGGACCCCATGCGGAGGGCGTGGTGGTGGTGGTGGTGGTGG 25518  
 25518  
 621 GAAAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 680  
 Db CGGAGCACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 25458  
 25458  
 681 AACCTCAAGTTTATTCAGCGGCTGAGGCAAGATCGTTTCTTCAACAATCTCTGG 740  
 Db GATCATGAGGAGTGGTGGGAGGAGCGCTGGCAGGACCGCCGCTGAGCTGGCGCAC 25398  
 25398  
 741 CGTGCACCTTCCGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 800  
 Db GAAACACGCTGCTGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 25338  
 25338  
 801 CGGAGAGATTAATGAGAACTGGGGGAGAGCTGCTCAAGCTTTTCTGCTGCGCTGG 860  
 Db GCGCGATCTGCTGCTGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 25278  
 25278  
 861 CGCATTTGAGCCCGGCTGCTGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 920  
 Db TCGCATCTTGGCGGAGCAACGCTGGGCTGGGCGGAGTGGGGGCGCTCGCTGCTGT 25218  
 25218  
 921 GATTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 977

25158 CATGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 25099  
 978 TGGCATTTCTGCTGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1037  
 Db 25098 CGGATGAGGCTTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 25039  
 1038 CGCGAATTTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1097  
 Db 25038 CGCGAATTTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 24978  
 24978  
 1098 GTCCAGCTTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1157  
 Db 24978 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 24919  
 24919  
 1218 GTCCGCAATTTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1277  
 Db 24858 GTCCCGCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 24739  
 24739  
 1278 GGAATCATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1337  
 Db 24798 GGAATCATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 24739  
 24739  
 1338 CGGATGCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1397  
 Db 24738 CGGATGCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 24679  
 24679  
 1398 CCTCTCGAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1457  
 Db 24678 CCTCTCGAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 24619  
 24619  
 1458 TGAAGCTTTAGC 1469  
 Db 24618 TGAAGCTTTAGC 24607

RESULT 12  
 CGLT/GC 3013 bp DNA linear BCT 17-FEB-1995  
 LOCUS C-glutamylcysteine synthetase and ORF.  
 DEFINITION X66112  
 ACCESSION X66112.1 GI:505580  
 VERSION X66112.1  
 KEYWORDS citrate synthase; glt gene.  
 SOURCE Corynebacterium glutamicum  
 ORGANISM Corynebacterium glutamicum  
 Bacteria; Actinobacteria; Actinobacteriales; Corynebacteriaceae; Corynebacterium.  
 REFERENCE 1 (bases 1 to 3013)  
 AUTHORS Elkmann, B.J., Thum-Schmitz, N., Eggeling, L., Luedtke, K.U. and  
 Sahm, H.  
 TITLE Nucleotide sequence, expression and transcriptional analysis of the  
 Corynebacterium glutamicum gltA gene encoding citrate synthase  
 Microbiol. 140, 1817-1828 (1994)  
 JOURNAL JOURNAL  
 DIRECT SUBMISSION  
 TITLE Submitted (15-MAY-1992) B. Elkmann, Institut f. Biotechnologie 1,  
 Forschungszentrum Juelich GmbH, I B T, Postfach 1913, 5170 Juelich,  
 FRG

FEATURES  
 source  
 1. .3013  
 /organism="Corynebacterium glutamicum"  
 /mol\_type="genomic DNA"  
 /strain="ATCC13032"  
 /db\_xref="taxon:1718"  
 prim\_transcript 717..>2151  
 gene 838..2151  
 /gene="glt"  
 838..2151  
 CDS

```
/gene="glt"
/SC number="4.1.3.7"
/codon start=1
/transl table=11
/product="citrinate (gl)-synthase"
/protein_id="CAA46902.1"
/db_xref="GI:505581"
/db_xref="SWISS-PROT:P42457"
/translation="MFERDIAVATNNKAVLHYGGEFENDITEASRNNNGVILGKLS
ENGILTFPDGVSSTGTSKITYIDGDAIILHVRGDIADIAENATFNNVSYILNGE
LTPPELHKFNEIRHHTLLDEDFKSQVVFPRDAHPMAATLSSVILSTYQDQNP
LDEADQDKATVRLMAKVPMLAAVAHRAKGAAPYMDNLNARENFLRMFGYPERY
EIDPINKALDKLLILHADHQNCSSTVRMIGSAQANMFISAGINALEHGEA
NOAVLEMLEDIKSNHGDAEFENKKNEDGVRLMGFGHRYKXNDPEALYKETAH
EILEHGGDDLIDIAIKLEIALADYFISRLKYPVDYTGILYVAMGFPDFTVL
PAIGRLPGWIAHRYBQLGAAGKINRPROVYGNESRKLVPEER"
2276..2632
/note="ORF"
putative binding protein"
/codon start=1
/transl table=11
/protein_id="CAA46903.1"
/db_xref="GI:50548"
/db_xref="SWISS-PROT:P42458"
/translation="MEKPOIELPVGPAPEDLVISDIIVGEGAAAPRGGEVHYVGV
FETGEFDSMDRGTSQFPLNGLIAGWQEGIPGMKVGGRQLTIPEAAVYGEBSGH
PLSGRTLVIIIDLSA"
BASE COUNT      716 a      818 c      799 g      680 t
ORIGIN
Query Match      25.3%; Score 372; DB 1; Length 3013;
Best Local Similarity 100.0%; Pred. No. 2,7e-66;
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1102 AAGCTTATGAGCGTCCCTCCCAACCGTTTGGTATTTAGTCAGGCGCTTGGCGG 1161
DB 3013 AAGCTTATGAGCGTCCCTCCCAACCGTTTGGTATTTAGTCAGGCGCTTGGCGG 2954
QY 1162 GCGGCGCTGCGATGTCGCCGCCGATCTGTCTACGCGCGGAAACCGTGGCTTCC 1221
DB 2953 GCGGCGCTGCGATGTCGCCGCCGATCTGTCTACGCGCGGAAACCGTGGCTTCC 2894
QY 1222 GCATTGCCACAGAGGCGCTCGGCATCTCTTCCGAGACCAACAGCGCGCGAA 1281
DB 2893 GCATTGCCACAGAGGCGCTCGGCATCTCTTCCGAGACCAACAGCGCGCGAA 2834
QY 1282 ATCATAGAGCGAACAAGCGGTGACGAGCGACATTTTAAGCCAAAGGCTTATCGACGG 1341
DB 2833 ATCATAGAGCGAACAAGCGGTGACGAGCGACATTTTAAGCCAAAGGCTTATCGACGG 2774
QY 1342 ATCGTCGCCGAAACCGAGACATTTGTTGAAGAAATCTTGGCAATACGCAACGCCCTC 1401
DB 2773 ATCGTCGCCGAAACCGAGACATTTGTTGAAGAAATCTTGGCAATACGCAACGCCCTC 2714
QY 1402 TCCGATTGGATTAACAATCCGAGAGAGGCGGACCGGACAGTCGCTTCAACGATTGAG 1461
DB 2713 TCCGATTGGATTAACAATCCGAGAGAGGCGGACCGGACAGTCGCTTCAACGATTGAG 2654
QY 1462 CGTTTACGCGAG 1473
DB 2653 CGTTTACGCGAG 2642
```

```
RESULT 13
AE006979/c      AE006979      17910 bp      DNA      linear      BCT 27-APR-2001
LOCUS           Mycobacterium tuberculosis CDC1551, section 65 of 280 of the
DEFINITION      complete genome.
ACCESSION       AE006979 AE000516
VERSION         AE006979.1 GI:13880475
KEYWORDS
SOURCE          Mycobacterium tuberculosis CDC1551
ORGANISM        Mycobacterium tuberculosis CDC1551
```

```
REFERENCE
AUTHORS
1 (bases 1 to 17910)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
Unpublished
2 (bases 1 to 17910)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 17910
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
145..1326
/gene="MT0918"
145..1326
/gene="MT0918"
/note="identified by Glimmer2; putative"
/codon start=1
/translation="hypothetical protein"
/protein_id="AAK45164.1"
/db_xref="GI:13880476"
/translation="MPSRATVOERSDSYPFCHNGRRPIMPKIYVSQSTRRLTSFV
GRKALNDVRLSDRLVTLTGPDMKSRSLAQIGAIHIEFTYGGWDCDLAVTD
RDQVYSIMNLALGVPQGLAIDTLVAVINDARVTLVLDHEHLIDCAALIDSLR
SCPRLITLSTEAIGAGELTVRPPLINDAIELVDRABRVRSPAIYADAVT
VGEICRRLDGVPLAIELAARDTSLPVEILIGNDRFIVAGAGNVRPQTCAT
VQSHALSGEPRLHRLAVPAGFPDLDGAQAVANDEDEGYTTLGRPAELVYKAT
VVENNRGRAGRLILYSVQYALEKLSSESGADAVLARYRHLKQNPQNVVAGSGVR
Y"
1358..2920
/gene="MT0919"
1358..2920
/gene="MT0919"
/note="identified by Glimmer2; putative"
/codon start=1
/translation="conserved hypothetical protein"
/protein_id="AAK45165.1"
/db_xref="GI:13880477"
/translation="MTAPPIAVERNTRSKVROQOEADVVALGRKPELLCVPERFAMD
LPMAADALFLMAETPRPLHVALVLSQPDNGTGRVLRVFSVAARQOVAPMWR
RPHRSITLIGOSMRTETEDVDLHYVLSAPRAGTAELVALVELHAGMIDRSPPL
MOVDLIEGLPGRCVAVYVHVALADGVSWRLORITVADPHOROMPTLWEVPOAS
VAKHTAPRGSRRPLTLAKGVIGARGVGWVRVADDTWRAPACGSGSLTLAAPTPL
NEPIAGARSVAGCFPIERLPOVAEHADATINDVYLVACGALRAYILSRALPEAPL
IAMPVSLDRTAVIDVFGQGNKIGTLMCSLATHLASVERLSAIRASMDGKAIA
GRSRNOALAMGALPAALMALGRVAPLRPPVITISNVPGQALYMGARLADAL
YLSAPVDGALNITSCGTNETFGTLGCRRAVPALSIITDQLAHELELLVGVSEAG
PSTRRLRIAGRR"
3093..4388
/gene="MT0920"
3093..4388
/gene="MT0920"
/note="similar to PID:2440108; identified by sequence
similarity; putative"
/codon start=1
```

```
/transl_table=11
/product="citrate synthase"
/protein_id="AAK4516.1"
/db_xref="GI:13880478"
/translacion="MADTDPTATLRYPGGEIDLOIVHATEGADGIALPPLAKGHTT
PDVGPAANAASSTITYIDGAGILRYGYPIDQIAEKSTFEVCYLIIYELPDTDO
LAQFPIORHPTMLHEDLKRFDGPPRAHMPVSVNALSAAYODALPMDNGQV
ELASTRIAKLPTIAAYAKKSGOPFLYPDNSLTVENFLRTPGPAEPYODAEV
VRALDMLFIHADHCONCSTVRLVGSRRALFTISGGTALMGPIAGCANVLE
MLEGIRDSGDVSEFRKYKREAGVKLMGEGHRKYKYDPRARIYKQADKIYAKLG
GDDSLGIKLEBEALTDYDFIERKLPNDPFTGLIYRALGFTPMFTVLFALGRL
PGLIHMREHMEHDEGSKIGRPQIYTGTYERDYVITIDAR"
/gene="MT0921"
/complement(4429. .6036)
/gene="MT0921"
/complement(4429. .6036)
/notes="similar to PID:1001724 PID:1001724, identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="oxireductase, putative"
/protein_id="AAK45167.1"
/db_xref="GI:13880479"
/translacion="MSDHRDFDVVVGGGNGLVAAAYLARAGLRVRLRLAQTCG
AAVSIQAFDGVVALSRYSYLSPRIADLDGAPVRLARPFSSTYPPATARRSG
LLIGTPBRAHILAIGAAPDAGFAFRRCRITVTAIMPILIEPIRTPEOARSI
VEYGEHAAMQAMVDEPDIHAIGAANDLRCVITADLIGTPAMHPSIMONI
CFLHVGCGTCVMVPIGMSGVTTSALATARRGAEITVTAQVFPALDPSTYHS
DGSDEHLVRGFPVLGVTAPVAVLSLGEPPAALAPGAQVNNVVRLLPRLRDSV
TPOQAFGTFHNETMSQDAAVSOASGRLEPDPICBAJCHSLTDPBILSLRUDAG
AQTIVFGHTPHSVGDTGELAERLTAVALINSVLAEPIDVLMTDQASKCIET
TTTLDLORTLGTGNIPIHGAISWPFADNDLDPAPQWGVATDHERIMLCGSGARR
GGAVSGIGSHNAAMVLAICASRRKSP"
/complement(6062. .6325)
/gene="MT0921.1"
/gene="MT0921.1"
/complement(6062. .6325)
/gene="MT0921.1"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45168.1"
/db_xref="GI:13880480"
/translacion="MGKGRKPTDSETHAHIRDLVAEKRLRAQLRHGISESEQQL
RRIEILDQCMLLQRRALRQTGDPREAVVRAPDQVEGTG"
/gene="MT0922"
/gene="MT0922"
/gene="MT0922"
/gene="MT0922"
/notes="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="OmpA family protein"
/protein_id="AAK45169.1"
/db_xref="GI:13880481"
/translacion="MASRAGLGGTATTDARTOKVRGSPRPWLIGAVVPIILIA
ICYGAFRRPQSTGPTGVLPITPTSTGASLSLSLSISGCTVTLIDFPBEAA
KAALMTALNGLAPGVNVIQIHDPVVRSLDFSABEPFTASVPIPDGLKVEEDIV
TLTGTPSBEHDAVKRAATSTWPMKLVNNIEVGOAPPGPAGPCADLISAINAV
TGGPIAFGNDGASLIPADYEILNRVADKLKACPDARVINGTDTGSEGINIPLSAQ
RAKIVADYVAVAGVDHIAITVGLGSVNPVLSNATPEGRAKNRREIVVN"
/gene="MT0923"
/gene="MT0923"
/gene="MT0923"
/gene="MT0923"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45170.1"
/db_xref="GI:13880482"
/translacion="MDFVIQWSCYLLAFLGSAVAVMVVVTISIKRASRDEGAAREASA
AETGAO"
```

```
gene
7575. .8105
/gene="MT0924"
7575. .8105
/gene="MT0924"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45171.1"
/db_xref="GI:13880483"
/translacion="WMEHVHMLAGLAFITGMVLTSTLMRPVPHQVLVKKSVGSSA
KSKPPIYAKPAVKSGETESPTAKTVATESAAEQIPVAGEPAEPTVAGEPAARI
PVVYAPYGGSSARAGAGSGSPQMWLVKRSBTRLYTTPEDPTVPTVAQWFEDEBS
AARAFPTPMRSTRRT"
/complement(8122. .9531)
/gene="MT0925"
/gene="MT0925"
/complement(8122. .9531)
/gene="MT0925"
/complement(8122. .9531)
/notes="similar to PID:2440113; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="sensor histidine kinase"
/protein_id="AAK45172.1"
/db_xref="GI:13880484"
/translacion="MALGCCIPSAESDCCVCSGACMNILSRIFARTPSLRTVVVA
TALCAALPVLIGTVVWVGITDRKERLDRIDRAGAFAPVPRGIDEIPSPNDOD
ALITVRGNWIKNSDITLPRQDYPADYVVRGVRVRYVTEGPTSVAGATYD
ATVATENNLSHERVLLICTFPAIAAVFAMLLAFAVAFEPQLAEQTSIDGDAFV
EVHASASLIEAEKRMQLORIMEONRTBALASABPAVSSHELRTPLTARTNL
EVALSTLDPDORKEVINLDYRITORSLEATLSALERLAQELSTSDHVPVDITDLD
RAHDAARIYDLDLVSPTCTIVGLPAGLRILAVNALANAVHGCATVLOLSAVS
SRACVEAIDNCSGVPEGRQYVEHFRSNGTASHSGSGLALVAQQLHSGTAS
LENSPLGABRLVLRPSPS"
/complement(9473. .10183)
/gene="MT0926"
/gene="MT0926"
/complement(9473. .10183)
/gene="MT0926"
/notes="similar to GP:2440114; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="DNA-binding response regulator"
/protein_id="AAK45173.1"
/db_xref="GI:13880485"
/translacion="MGGMDTGVTSPRVLVVDDSDVLASLERGLRLGFEVATAVDGA
Query Match 21.2%; Score 312.6; DB 1; Length 17910;
Best Local Similarity 54.0%; Pred. No. 3.7e-54;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;
```

QY	65	ACTGGCTTTTGACCCCTGACAGCTTCATTTTGGAAATGAACCTCCCAATATACAAAC	124
DB	11770	ACGGCGTCTAGACCGGGGATCTTTCGACAGCTGGATAGCAACCGCTGCGGTCGCGG	11711
QY	125	TCAATCAAGGCTATAGCAGAGCCTTGGAGCGGGCTCGAAGCAAGCCAAATGCGATGAT	184
DB	11710	TAGCCGACTCTATAGCGCGGAGCTGCGCGCGCTCGGCGCGCAACCGCGCGACGAAT	11651
QY	185	CGTAATTAATCTGGAAGAGCAGCTGGAAGGCAATTCGCGTACCGGTATTTTGTCCGATT	244
DB	11650	CGGTGACAGACCGGAGGAGGAGCGGTATTCGGGCGCGCGGTCGCGTGTGAGT	11591
QY	245	TTTCTCTCTCGCGGCTCTTTGGGACAGGTCGCGTGGTCGATCATGAAGCGATTC	304
DB	11590	TGCACTTCTGCGGCGCTCGATTGGGTCGAGCGGCGCAACGATACCCGCCCGCGC	11531
QY	305	ACCGGCGACAGAGCTGAATCTCCCACTGCTGCTCCCTCGCTTCGCGTGTGCGGCA	364
DB	11530	AGCGGCGACCGCGGAGCGGCTGCGCTACCTGCGGTCAACGCTCGGAGGACCGCGCA	11471
QY	365	TGCAGGAAGACATCGAGCTTTTGTCAATGATGTGTCCATTAACCGCGCTGTGAGCGTC	424

Db	11470	TCGACGAAGGACGCGTCCGTTCTTCGACGATGATGAAAGATGCCTGCGGCGCATCCAGCTGC	11411
QY	425	ACCGGAGAGCCCATTTGCGTTCTCGTGCTATTTTCCGCAATCCACGATGGGTGGCGCCA	484
Db	11410	ACAAACAGACGCGCCCTGCCTCACTGGCTCTATTTGGCCCATCCGACCAACGGGTGGAGTTT	11351
QY	485	TGGCCTCTGAGGGATTCACTGAGGCATCTCATTTTGGGGAACCCGAGCGGCACATAGGTT	544
Db	11350	TGCGCTCTGGGGGCTCGTGGGGCATCTCACCGCTCCGAGCCGGGGGCCCTTATAGGCT	11291
QY	545	TCCTGGTCTCGCGTGTGAGTTAACCATGCGGCATGCGCTTCACAGCGTGTGACG	604
Db	11290	TTCTGGGACCACGGGTCTATAGTTGCTCTATGGCGACCCCTTCCATCCGGGCTCAAA	11231
QY	605	AGGGGAGAAATTTGGTGAATACTGTGTGATGTATGAAATTTGTGTCCCACTCTAAATTC	664
Db	11230	CCGCGGAACATCTACGGCGGAGATGGATCATCCGACGCGCTGTTCATCTGACCTGGCTAC	11171
QY	665	GTGCAGCGGTTGGCAAAAACCTCAAGGTTATTCAGCGGATAGAGGCAACGATCGTTTTT	724
Db	11170	GACCGATGCTGGATCTGTGCTGTGACGGTGTCTCATCCAGCTCCCGAACCGCTTCGGCAC	11111
QY	725	CTCCAAACAATCTCTGGCGTGCA---CTTCGGTATGAGGCGGATTCGCGTTCTCGTG	781
Db	11110	CGCAGACGCCCGCGCCCGTACCCGATGTGCCACGTGGACTCGGTGTGACATCCGCGC	11051
QY	782	ACCGCGAGAGCGCTGGAATCGGGGAGATTTATGAAACGTTGGGGGCAAGCTGTCTAAC	841
Db	11050	GCGCGAGCCGGCGGGGCTCAGGACGCTACTGTCACACGCGCCACCTGACCCGGGTCTTGT	10991
QY	842	TTTCTGTGTGCGCTGTGCGCATTTGAGCCCGCGTGTGCGGCTTGGCCCTGAGCGCGATG	901
Db	10990	TGTCAAGAACCGATCAAGGGGAGGCG---GGAACACGCTGCTGGCGCTGGCCCGTTTG	10934
QY	902	GGGGCCGGCCCTGTGTCTGATTTGGGAG-----GATCGCGCTTCAAGC	946
Db	10933	GCGGCACAAACCAACGTGTGTCTCTCGGCGAGAAAGGACGATGAGCGCGGGGAAAGCATG	10874
QY	947	TTGGGCGCGCAGAGACTCGTTTTTTCGCGTGTGGCATTTTCGTGGCGCGAGCTTAAAC	1006
Db	10873	TGGGGCCCGCTGTGCTTACGGAGACCCGAGCGGGATGCGCTCGCCCGCGAGCTGTGCC	10814
QY	1007	TGCCGATCTGTTCATCATGACACCTCGGCGCCGAAATTTGTGACAGGCGGCTGAGAGAC	1066
Db	10813	TGCCGCTGTGTGTGATTTGACGCGGACCGAACCCGCGTTGTGTGGCGGACCGAACAGG	10754
QY	1067	TGGGCATTCGACGTTCCATTGGCGGACCTTTGTTCAAAGCTTATTCACGCTCCCTCCCA	1126
Db	10753	GCGGGCTGGCGCGCAATGCGCGATTGCTCTGGCCGAGCTGTCAAGCTGAAATCCCGCA	10694
QY	1127	CCGTTTGGGTCATTATTTGATCAGGGCGGTTGGAGGTGGCGCGCTGGCATGTGCGCCGCG	1186
Db	10693	CCGTGTGATCTGTCTGTGGGCAAGGCAAGCGGCGCGCGCTGGGGAATGTTGCCCGCGG	10634
QY	1187	ATCTGTGTCAGCGGCGGAAACGCGTGCTGTTCGCAATTGCCACGAGAGGCGCCTCGG	1246
Db	10633	ACCGGGTGTGGCGCGCATCTCAAGCGTGGCTGTGCGCCTTGTGCTCCGAAGAGCCAGG	10574
QY	1247	CCATCTCTTCCGCGAACCAACACGCGCGGAAATTCATAGAGGCAAGGCGTGTGACG	1306
Db	10573	CGATCTGTCTCCAGACACTGTCTATAGCGCGCAACTCTGCTGCGGCCCAAGGCATTCGGT	10514
QY	1307	CGCAGCAGCTTTTAAGCAGAGGCTTATCGACGGGATGTGTGCGCGAAACGAGCACTTTG	1366
Db	10513	CGGCGCAGACTAGGAATCGGGGATTTGTGACACCATCTGTGCGCGAGTACCCCGACGCG	10454
QY	1367	TTTGAAGA 1373	
Db	10453	CAGACGA 10447	

RESULT 14  
MTCY31/c

DEFINITION	LOCUS	37630 bp	DNA	linear	BCT 03-AUG-2001
MYCOBACTERIUM TUBERCULOSIS H37Rv complete genome; segment 41/162.	MTVCY31				
ACCESSION	273101	AL123456			
VERSION	273101.1	GI:3261565			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
MEDLINE					
PUBMED					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
COMMENT					



```
/db_xref="SWISS-PROT:Q10536"
/translation="MRIGVCTTPDPAQAAVEAAGARDELAGEPAPSLAVLLGSRRAH
TDRADVLAVALQMDIPALVGCIAQAIAGHNEIDEPAAVVMVLAAGLAETPOLDR
VRTSGALITGRFPRTARADLLHLLPPTTFSNLLIEHNTDLPGTAVGVSGGR
RGGDRLFRDHVLTSGVGVRLPKRGVPVSGCRPIGPIYIVTADGILLITLGG
RPLRLREIVELSPDERALVSHLQIGIVDEHIAAPGGDFVIRGLGADPTGS
EIDVDVVGATMOQOVDRADAGADDLRLTERAARLRGAAAGALLFTCNCRGRME
GAVHDASTIELLGGIPLAGEFAAGEIGPIAGRNALHGFTSMALLPVDME"
complement(1314, .1802)
/gene="Rv0875c"
complement(1314, .1802)
/note="Rv0875c"
/gene="Rv0875c", (MTCY31.03c), len: 162, unknown, possibly
exported protein, highly similar to MCB57.11 Mycobacterium
leprae cosmid B57, 033056 HYPOTHEICAL 16.1 KD PROTEIN
(162aa). FASTA scores: opt: 789 z-score: 963.4 E(): 0;
71.4% identity in 161 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0875c"
/protein_id="CAA97382.1"
/db_xref="GI:1314012"
/db_xref="SWISS-PROT:Q10537"
/translation="MKRGVATLPVILVILLSVAGAAMLVLRGHGPOQPEISAYSHG
HLTRVGYLYCNVDLDCCQPOAGELPVSESRYPQLSVEPVISRAPMRLLOVODP
ANTSTLPRPDRLAVTIPTVDPQGRGLGIIVQLTLVVDHSGELRVRVPAHMSVRL
IP"
complement(1799, .3445)
/gene="Rv0876c"
complement(1799, .3445)
/note="Rv0876c"
/gene="Rv0876c", (MTCY31.04c), len: 548, unknown, possible
membrane protein, highly similar to MCB57.12
Mycobacterium leprae cosmid B57 033057 HYPOTHEICAL 61.7 KD
PROTEIN (579aa). FASTA scores: opt: 2850 z-score: 2863.4
E(): 0; 81.0% identity in 568 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0876c"
/protein_id="CAA97383.1"
/db_xref="GI:1314013"
/db_xref="SWISS-PROT:Q10564"
/translation="MAPTGGRTGRNGSVNGHGMANYPPDDANYSRRPPMPGSANR
YLPGLGDEPERSRVPRTTRRGRITVYTRAARSBMGRMVLVLRATATAGAD
KSGLTALPMWANAIVASAMVALANTLFPAAAGSGSKSPVALYLTITIPRATIR
LIGPALDRLOHGRVVALALSGLTALAVLIMYDGTGSPSWLTPCALAMVIES
KSFVLSAATVPVPTIDLVANSRLVFLGLGTAGALAAVGFVCTHLPQLP
GALFVVAITTAGISLSKRIPIRWEVTEGSEVATLSYHRDRGLRRRPEEVKNGGT
LRQPLGRNITISMGNCITIKVWGFLFLYPAPVAKAHANGVQLGMGLGAAVAG
NPAQFTSARLQIGRPALVAVCTVAVVIAATAAVAGSLATATATLITGSSALIAK
ASDASLQHDLPBESRAGSFGSESTLQATWLTGCAVGVLYTTELWGTAVSALLIL
GLAQITVSRGSLIPGLGGRNRPVMAEDSTTKRGAAVAQ"
complement(1457, .3465)
/note="gagstgac, potential rbs upstream of Rv0876c"
3572..3575
/note="gagc potential rbs upstream of Rv0877"
3583..4371
/gene="Rv0877"
3583..4371
/gene="Rv0877"
/note="Rv0877"
/gene="Rv0877", (MTCY31.05), len: 262, unknown,
equivalent to MCB57.13 Mycobacterium leprae cosmid B57
033058 HYPOTHEICAL 28.2 KD PROTEIN (269 aa; E(): 0; 80.5%
identity in 257aa overlap.)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0877"
/protein_id="CAA97384.1"
/db_xref="GI:1314015"
/db_xref="SWISS-PROT:Q10539"
/translation="MTGPTBESAVATVADMEGLAVALRGADQARAAVVESGPEAV
GDYLGVSIEDGAATSPRIALHFGYQGMQMAVAVASYSAGDAATISEVVLVGPALIL
APDWVWEGRVAPGDLSPGDLAPAKDDPRLVPGVYTAGADAVDTAETALIGRWVW
```

```
SAWRAOASORWHDGDYPGSARASTKRVCRDQCFPLPLAGSLGAMFGVCNGLSAD
GHVVDROYGCGASHDDTAPAGSGSPPIVEPYDDVDLIDIEKPAES"
complement(4380, .5711)
/gene="PPE"
complement(4380, .5711)
/gene="PPE"
/note="Rv0878c", (MTCY31.06c), len: 443, unknown, similar
to YH56 MYCT0 P42611 hypothetical 50.6 kd protein in hsp55
3' region. (517 aa). FASTA scores: opt: 1044, z-score:
954.9, E(): 0; (47.4% identity in 397 aa overlap). Similar
also to MTW014_3; MTC165_2, MTC198_24, MTC1307_23,
MTCY48_17, MTW004_5, MTW004_3 etc."
/codon_start=1
/transl_table=11
/product="PPE"
/protein_id="CAA97385.1"
/db_xref="GI:1314016"
/db_xref="SWISS-PROT:Q10540"
/translation="MNFVLPPEVNSARIYAGAPAPMLAAVAMDGLAELGMAAS
FSLISGLTAGPGSGAPPAAMAAAYLSLWLNATARAEGAAARAAVAYEA
ARAATAPALVAARNOLLISLVSNLPGMLPATAETASVEQLMADVAAGVHGG
ASTVASQLTWQQLSTLPPVTPAARGAVPAALAIPLAGVENTGVFLGINTG
NNVSGNTGDYNEIGNIGNANIGNANIGNAGAFENFNGNDNTNFGSN
AGFLNTSGNBSGNLGFNAADNTGNSGDTNFGNSGDLNTGSPVTOGVAN
SGFNTGTGSGFNSSGNSGSPONLNGSSGFNADSTSSGFONACTALTASSTWA
DSPRAWETIRAPSRLOVWRTATTAECISIVIISRVSTGAPPKKVGNGS"
complement(5719, .5721)
/note="yga (or aeg) potential rbs upstream of Rv0878c"
complement(5989, .6264)
/gene="Rv0879c"
complement(5989, .6264)
/note="Rv0879c"
/gene="Rv0879c", (MTCY31.07c), len: 91, unknown, equivalent
to MCB57.14 Mycobacterium leprae cosmid B57 033059 HYPOTHEICAL 9.8
KD PROTEIN (91 aa), E(): 1.2e-25, 76.9% identity in 91 aa
overlap."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0879c"
/protein_id="CAA97386.1"
/db_xref="GI:1314017"
/db_xref="SWISS-PROT:Q10541"
/translation="MSVENSGQIREPPLPVLLEWVPIAVGALMLVAVAAFPVPG
LASRPVTVAGLATGLLGTITIFWQLAARGARQAQAGETYLDPK"
6442..6873
/gene="Rv0880"
6442..6873
/gene="Rv0880"
/note="Rv0880"
/note="Rv0880", (MTCY31.08), len: 143, unknown, equivalent
to MCB57.15 Mycobacterium leprae cosmid B57 (143 aa)
033060 HYPOTHEICAL 15.8 KD PROTEIN. FASTA scores: opt:
818 z-score: 980.3 E(): 0; 89.5% identity in 143 aa
overlap."
/codon_start=1
```

Query Match	21.2%; Score 312.6; DB 1; Length 37630;
Best Local Similarity	54.0%; Pred. No. 3.4e-54;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;	
QY	65 ACTCGTTTGGACCTGACGCTTCATTTCTTGAATGAACCTCCCAATATGACAAC 124
DB	35657 ACGGCGTCTAGACCGGGGATCTTTCCTCACTGAGATAGCAGCCCTGCGGTGCGG 35598
QY	125 TCAATCAAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGCCAAATGCGATGAT 184
DB	35597 TAGCGACTCTATAGCGCGGAGCTGCGCGCTCGGCGGCGACCGCGCGGAGCAT 35538
QY	185 CGGTAAATTTCTGGAAGGCACTGGAAGGCAATTCGGTGAACCGCTTATTTTGTCCGATT 244
DB	35537 CGGTGACAGACCGGTGAGGAGCGGTATTCGGGCGCGGCGGTGCGGTGCGCTGATG 35478
QY	245 TTTCCTTCCTGCGGCGGTCTTTTGGGACCGGTCGGCGGTGCGCATATGAAGCGATTC 304



Db	354177	TCGACCTTCTGGGGGGCTCGATTGGGGGTGGAGAGGGCGGAACGGATACCGGCGCGGTGG	354118
QY	305	ACCGGGCCACAGAGCTGAAATCCCACTGCTGGTCTCCCTGCTTCGGGTGGTCGGCA	364
Db	35417	AGCGGGCCACCGCGGAGCGGCTGGCGGTACTGGCGGTACCAAGCTCGGGAGGCAACCGCA	35358
QY	365	TGCAGGAAGAACAATCGAGCTTTTGTCAATGATGGGTGCATATACCGGGGCGTGTCAAGTTC	424
Db	35357	TGCAGGAAGGACAGGTGCGCTTTTGTGCAGATGTGAAGATCGCTGGCGGCATCACTGG	35298
QY	425	ACCGGAGGCGCATTTGGTCCGTTCTGTGGTGAATTTTGGCCAAATCCACGATGGGTGGCGCA	484
Db	35297	ACAACAGAGGCGCGCTGGCTCCACTGGTCTATTTGGCGCATTCGACCAAGGGGTGAGTTT	35238
QY	485	TGCGCTGTGGGGTTTCATCTGGGCATCTCACTTTTGGCGAACCGGCGCGCAGATAGTT	544
Db	35237	TGGGTGTGGGGCTTCGCTGGGCGCATCTCAACCGTCGCGAGCGGGCGCCGTATGGCT	35178
QY	545	TCCGTGGGTCCTCGGGTGGAGTTAACATCGGGCAATGGCGTTCCAGACGGTGTACAC	604
Db	35177	TTCTGGGACACGGGGTCTATGATGTCTATGGCGACCCCTTCCATCGGGGTCAAA	35118
QY	605	AGCGGGAATTTTGGTGAATACTGGTGTGATATGAGAAATTTGTGCGCACTCAATTCG	664
Db	35117	CCGCGAGAAATCTACGCGCGGATGGATTCATCGACGGCGCTGTGGCACTGGACCGGCTAC	35058
QY	665	GTGCACGGTGGGCAAAACCCTCAAGGTATTCAGCGGTAGAGGCAACGATCGTTTTT	724
Db	35057	GACCGATGTGGATCTGTGGTGTGACGGTGTCAATCGACGCTCCGGAACGGCTTCGGGAC	34998
QY	725	CTCCAACAATCTCTGGGTGGCA---CTTCCGGTATGAGAGGGGATTTGGCGCTTCGTG	781
Db	34997	CGCAGACGCCCGGCGCCGTATCCGATGTGGCCACGTGGGACTGTGGTGTGGCAATCCGCGC	34938
QY	782	ACCGCGAGGCGCTGGAATCGGGGAGATTAATGGAACGTTGGGGGCGAGACGTGTCAACG	841
Db	34937	GGCGCGACCGGCGGGGCGTCAAGGACCTTACTGCGACACGGCGCCACCGACGGGTATTGT	34878
QY	842	TTTCTGTGGCGGTGTGGCGCATTTGAGCCCGGCTGTGCGGTTCGCTTGGCGCGCATG	901
Db	34877	TGTCAAGAACGATCAAGGCGGAACG---GCGAACACGTGTGGCGGTGGCCCGCTTGG	34821
QY	902	GGGGCCGGGCGGTGGTGTCTGATTGGGAG-----GATCGCGCTTTACGC	946
Db	34820	GCGGCGAACCCACGCTGTGTCTCGGCGCAGGAAGGCGAGTAGGCGCGGGGGAAGCACG	34761
QY	947	TTGGGCGCGAGAGACTCGTTTTCGGCGGTGTGGGCAATTTGGCTGGGCGCGAGCTAAAC	1006
Db	34760	TGGGCGCCGCTGCGCTTACGGAAGCCGGAAGCGGGATGTGCGCTCGCGCGCACTGTGCC	34701
QY	1007	TGCCGATCGTGTTCATCATCGACACTTCGCGCGCGCAATTTGTGCGAGGGCGCTGAGAGC	1066
Db	34700	TGCGGCTGTGTGTGTGATTTGAGCGCGGCGGACCCGGTGTGTGGCGCGCAACGAG	34641
QY	1067	TGCGCATCGCAAGTTCATTTGCGCGCACTTTGTTCAAAGCTTAATGACGCTCCCTCCCA	1126
Db	34640	GCGGGCTGGCGCGCAATGCGGATTTGCTTGGCGGAGCTGTCAAGCTGTATACCCCGA	34581
QY	1127	CCGTTTGGGTCATTATTTGTCAAGGCGGTGGGCGGTGGCGCGGCGCATGCTGCCGCGG	1186
Db	34580	CCGTGTGATCTGTCTGTGGGCGAAGGACAGCGGCGGGCGCGGGGCTGGGAGATGTTCGCCGCG	34521
QY	1187	ATCTGTCTACGCGGCGGAAGACGCGTGTCTGTCCGATTTGCCACAGAGGGCGCTTCGG	1246
Db	34520	ACCGGGTGTGGCGCGCACTCCAGCGGTGTGGCGGCGCTTGTGCTCCGGAAGGAGCAAGG	34461
QY	1247	CCATCTCTTTCGGGACACCAACACGCGCGCGGAATCATGAGCAACAAGCGTGTACAG	1306
Db	34460	CGATCGTGTCCGAGACATGTCTCATTCGCCCGCAACTCGCTGTGCGCCCAAGGATCTCGGT	34401
QY	1307	CGCAGCGACTTTTAAGCCAAAGGGCTTAATGACGGGATTCGTGCGCGGAACAGAGCACTTTG	1366
Db	34400	CGGCGGACTTACTAATCGGGGATTTGTGCACACCATTCGTGCGGGGTACCCCGAGCGCG	34341

Qy	1367	TTGAGA	1373
Db	34340	CAGACGA	34334

RESULT	15
EX248337/c	
LOCUS	327650 bp DNA linear BCT 11-JUN-2003
DEFINITION	Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment
ACCESSION	BX248337
VERSION	4/14.
KEYWORDS	BX248337 BX248333
SOURCE	BX248337.1 GI:31617663
ORGANISM	complete genome.
	Mycobacterium bovis subsp. bovis AF2122/97
	Mycobacterium bovis subsp. bovis AF2122/97

## REFERENCE

**AUTHORS**  
Garnier, J., Eigmeleir, K., Camus, J.-C., Medina, N., Mansoor, H.,  
Pyror, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempé, C., Simon, S.,  
Harris, B., Atkin, R., Doggett, J., Mayes, R., Keeling, L.,  
Wheeler, P. R., Parkhill, J., Barrell, B. G., Cole, S. T., Gordon, S. V. and  
Hewison, G.

**TITLE** The complete genome sequence of *Mycobacterium bovis*  
**JOURNAL** online Publication  
**REMARK** PNAS 10.1073/pnas.110426100 ( Microbiology )  
**REFERENCE** 2 (pages 1 to 327650)  
**AUTHORS** Garnier, T.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (24-MAR-2003) Garnier T., Unice de Genetitic

Direct Submission  
Submitted (24-MAR-2003) Garnier T., Unité de Génétique Moléculaire  
Bactérienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex  
15, France, e-mail:tgarnier@pasteur.fr submitted on behalf of the  
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary  
Laboratories Agency Weybridge, Woodham lane, New Haw, Addlestone,  
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA, UK. PR4 Amelcomme Genopole, Institut  
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.  
Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28  
rue du Docteur Roux, 75724 Paris Cedex 15, France

**FEATURES**  
**SOURCE**

**Source**

gene  
CDS

```

/locus_tag="BD0912"
/note="M090912, -, len: 490 aa. Equivalent to Rv0888, len: 490 aa, from Mycobacterium tuberculosis strain H37Rv, (100.0% identity in 490 aa overlap). Probable exported protein. Equivalent to Ark45157.1 from Mycobacterium tuberculosis strain CDC1551 (507 aa) but shorter 17 aa. Contains possible N-terminal signal sequence."
/codon_start=1
/translate_table=1
/product="PROBABLE EXPORTED PROTEIN"
/protein_id="CAD93773.1"
/db_xref="GI:31617664"
/translation="MDYAKRIQGVGALAVLVGVAATVTHATGSAPTDPSSTSP
VDCSPILGSSASLIAIPGASVPQVGRVQYDPOSITDDLLNATIDELAAVRLGVPII
ENRTPIYAPQOVSVPERGTVGVPFDCPDGKRMIFAREAKPGCPQHGITVDTOR
TASFIATADPGFVGDTGVNVSDDTSLPHVDFGPHGDVAIVVAFGNPT
DTISGFSMLITVNIAGLPEPLSAILPRFFYIEIKRNLAAVYVAVNOEFAHQEPI
KSKMPSQTPPEPPTLIMPIGVPSDLGATLSEFKTADROVTWCTSPNCTLKGF
TSSOMLPGDITVDVYVNLHTNGCGTPTANLAQVYAIQONSAGAAVITGPFNARY
SDOSALLQFAOVNGLTDAWVVOYEHGPTTPPAPFCMVNNECEGLDKIFIRSQGVTL
QAVSYENEARPKFNSKGEPLSDHSPAVVGFFHYADNVAAVR"
complement(2205..3326)
/gene="ctaA"

```

```
CDS
/locus_tag="Mb0913c"
complement(12205..3326)
/gene="cItA"
/locus_tag="Mb0913c"
/EC_number="2.3.3.1"
/notes="Mb0913c, cItA, len: 373 aa. Equivalent to Rv0889c,
len: 373 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 373 aa overlap). Probable cItA
(alternate gene name: gltA), citrate synthase 2 (EC
4.1.3.7), highly similar to others e.g.
CAB95899.1|AL359988 putative citrate synthase from
Streptomyces coelicolor (387 aa); P39119|CISY_BACSU
citrate synthase II from Bacillus subtilis (366 aa), FASTA
scores: opt: 586, E(): 5.8e-30, (33.8% identity in 367 aa
overlap), etc. Also similar to Rv0896|MTCY31.24 from
Mycobacterium tuberculosis (29.2% identity in 274 aa
overlap) and Rv1131. Contains P500480 Citrate synthase
signature. BELONGS TO THE CITRATE SYNTHASE FAMILY."
/codon_start=1
/transl_table=1
/product="PROBABLE CITRATE SYNTHASE II CITA"
/protein_id="CAD93774.1"
/db_xref="GI:31617665"
/translation="MTVVBNFVPGLDGVAFTEIAPEDKQGLARYGDIEDLVY
ORVTEGVWALLVDGNFGSGLPAPPEPLPHSGDVRVDVAGLAMLPIGVAPLID
IDDPARQOOLARASVMAVSVAOSAGIYOPAPORITDECSVTARFMRWQSPDP
RHIAIDAVWYSAEHGWNASTFRPARIASGADVAAALSGAIGMSRPLIGCAPRY
LPHIDEVRAGDASVVKGIIDRGKLMFGHRYRADPRAVIRAAEELGARIE
VAVAEQALSELREPRDRAIETIVERMAVVIDFAVPMMPMAMFTCGRTGWCA
HLEQRKGLREPRSAIYVGPSPESVDMERVLTIA"
complement(3413..6061)
/locus_tag="Mb0914c"
complement(3413..6061)
/locus_tag="Mb0914c"
/notes="Mb0914c", len: 882 aa. Equivalent to Rv0890c,
len: 882 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.9% identity in 882 aa overlap). Probable
transcriptional regulatory protein, LuxR family, highly
similar (but shorter 238 aa in N-terminus) to
NP_302202.1|NC_002677 possible transcriptional regulator
from Mycobacterium leprae (1106 aa). Also highly similar
(generally in part) to others e.g. T50568 probable
multi-domain regulatory protein from Streptomyces
coelicolor (1334 aa); P10957|NARL_ECOLI nitrate/nitrite
response regulator protein from Escherichia coli (216 aa),
FASTA scores: opt: 193, E(): 6e-06, (37.4% identity in 99
aa overlap); etc. Also highly similar to others from
Mycobacterium tuberculosis e.g. MTCY02B10.22, MTW008.44,
MTV036.21, and MTCY31.24. Contains P500017 ATP/GTP-binding
site motif A (P-loop)_P800622 Bacterial regulatory
proteins, luxR family, signature, and probable helix-turn
helix motif from aa 836 to 857 (Score 1559, +4.50 SD).
BELONGS TO THE LUXR/DHPA FAMILY OF TRANSCRIPTIONAL
REGULATORS."
/codon_start=1
/transl_table=1
/product="PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN
(PROBABLY LUXR-FAMILY)"
/protein_id="CAD93775.1"
/db_xref="GI:31617666"
/translation="MRALLAONRLVTLCTGGVGVKTRLAIOIASSELBDLCFVDLA
PITSGVIAATARAAGVLPDQGRSTMSLSRPTGCRPMILVINDCHELLDACAAVY
ELIACPELTILATSRPIGMAGEITRVPSMSTIDDAVELPADRASVVGFTTAN
NAAAVGEICRLDGLPIAIFRAAIVRSMSPLEIADGDDCFRLLAGVVRVAPQOT
LRASIDMSHALITEQILFRRLAPFVGPIAAVRAVASDIDPEFVLQTLVLD
KSLVADDCQGRTRYLETIRVALYEVLGSGEADVARRADYDTLAAALNTPAD
DHORLVAREITEIDNLRAPAFMSRENHITALQIASLQIIFGRRALRGSLFMS
ILDEORFRLAVSTAVRARALADKRMSTWLTATSPVGTDTIILAQOALLAARWPS
AAVRAITACGSSGYNAEAPYFAEITDARAIIDKMTTCQILHMGVETCISGDP
NALPRAAEERDLADITIGDRFVSRRCSIMLSLAWMANLLEALSELITAEVANS
DVPTKVLGLYQAOVLVAGSAANAIAGACIIAATEIGVYOGIGVAAVYALAA
DVTAALEASDARPLIRAQPOVTHWQVMDGPHAQTVAIIPGFNFEMMLGPSVQNFPL
RMVALTRAVATARGPELARDADHAALACGAEHLITQGMPEMALLAGLAGVSH
```

```
SEGRVLGAAALROQTRQVRFKIMDAGYQASVTLAEAMGDEDFDEAAGAALSTD
EALIAORGERRRRPARMGSLPTRERODVRLVSELSKDIARLFFVSPRTVQTHL
THYAKTGLASRVOLVDEARBSPS"
complement(6063..6920)
/locus_tag="Mb0915c"
complement(6063..6920)
/locus_tag="Mb0915c"
/notes="Mb0915c", len: 285 aa. Equivalent to Rv0891c,
len: 285 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.6% identity in 285 aa overlap). Possible
transcriptional regulator, highly similar in N-terminus to
NP_302202.1|NC_002677 possible transcriptional regulator
from Mycobacterium leprae (1106 aa). Also highly similar
to several Mycobacterium tuberculosis putative
transcriptional regulators e.g. Q102|MTCY02B10.22
PROBABLE TRANSCRIPTIONAL REGULATORY PROTEIN (1159 aa),
FASTA scores: opt: 702, E(): 8.3e-40, (50.6% identity in
247 aa overlap); MTW036.21, MTW008.44; MTCY02B10.23. Also
shows similarity with several adenylate cyclases and
hydrolases from other organisms."
/codon_start=1
/transl_table=1
/product="POSSIBLE TRANSCRIPTIONAL REGULATORY PROTEIN"
/protein_id="CAD93776.1"
/db_xref="GI:31617667"
/translation="MLEFVAVNSLPNIDIDHAILRGEDHPPTCAKCVARGISALGS
LDRBYSLRVCYAAPMPVGRCEVPRRRVLIANQGDVSLPTPGVTLLIAYVEST
HLMQCEEDVATIAIHDHTVSEAITTHGVOYKRYREGSPFAAFTRASDAACALD
LQRTSLAPILRLIGLTHGEVOLDELVTGVTIRTAALRLAAGGVLSAAAGTDLT
GRLPADMLVLDGRHPLRLGPRBEMWQLCPIDIREKPLRTAKSSPTSILPAQFTT
FVGRRAQIS"
7318..8805
/locus_tag="Mb0916"
7318..8805
/locus_tag="Mb0916"
/locus_tag="Mb0916"
/EC_number="1.14.-.-"
/notes="Mb0916", len: 495 aa. Equivalent to Rv0892, len:
495 aa, from Mycobacterium tuberculosis strain H37Rv,
(100.0% identity in 495 aa overlap). Probable
monoxygenase (EC 1.14.-.-), highly similar to others e.g.
NP_250787.1|NC_002516 probable flavin-binding
monoxygenase from Pseudomonas aeruginosa (491 aa);
CAB95668.1|AL132674 monoxygenase from Streptomyces
coelicolor (519 aa); P12015|CWO_ACS cyclohexanone
monoxygenase from Acinetobacter sp. (542 aa), FASTA
scores: opt: 489, E(): 6.8e-26, (30.3% identity in 492 aa
overlap); etc. Also highly similar to Rv0565c, Rv354c,
Rv3083, etc from Mycobacterium tuberculosis. Has
hydrophobic stretch at N-terminus."
/codon_start=1
/transl_table=1
/product="PROBABLE MONOOXYGENASE"
/protein_id="CAD93777.1"
/db_xref="GI:31617668"
/translation="MTCRCPTVAVVGAGSGMCVAITLISAGITDVCIEKADVGT
WRDVTYGLTCDVPSRLYOYFAKNPMWTFMRGGEIOOYLGIERYGRLRRIRG
ATVYSAPFDGRVLTPTDGSSTVDLFLSATGVLHPRIPPLAGLDPEFGTVFHSAR
WDITVPLGRRIAVIGTSGTVQVGLAVAGKVTMFGTAVQVLTMPMPRRSKLAR
VFHRAFPCLSLAVKAYSLETFEVALSNPGJHRKLVGACRASLERVDPRLRLAL
TPDIYEPKRLVMSGFFYRAIQRDDVELVTAGIDHVEHRIIVDDGLVHSDVIVLAT
GPDYHAFPRFMQITGRGRIIDVMDQGFPAQTVAIIPGFNFEMMLGPSVQNFPL
TVAESGAHHIVQMIKRWHRGEPDTEPKSAATEAVNTVLRAMPNTVWTGDSWVL
```

Query Match 21.2%; Score 312.6; DB 1; Length 327650;  
Best local similarity 54.0%; Pred. No. 2.e-54;  
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

```
Qy 65 ACTGGTTTGGACCCGACAGCTTCTTCTTGGAAATGAAACCTCCCAATATGACAAACC 124
Db 21614 ACGGGGCTAGACCGGGGATCTTTCTGCTAGCTGGGATACGAGCCCTGCGGTGCGG 21555
Qy 125 TCAATCAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGCCAAATGCGATGAAT 184
```

Db 21554 TAGCCACTCTATGCGCGGAGCTGGCCGCGCTCGGGCGGCAACCGGCGGAGCAAT 21495  
Qy 185 CGGTATTAATCTGAGAAAGGSCACCGTGAAGGCAATCCGGTAGCCGTATTATTTGTCCGATT 244  
Db 21494 CGGTGCAACCGGTGAGGAGCCGTAATTCGGGCGCGGGTGGCCGTGGTGGTGGTGGTGGT 21435  
Qy 245 TTTCTCTCTCGCGGTTCTTTTGGGCAAGGTGCGGTGCGGTGCGATCATGAAAGGCGATTTC 304  
Db 21434 TCGACTCTCTGGGCGGTGATTTGGGGTGGCAGCGGCGGAGCATGACCGCGCGCGTGG 21375  
Qy 305 ACCGGCCACAGAGCTGAAATCTCCACTGCTGTCTCCCTGCTCTTCGGTGGTGGCGCA 364  
Db 21374 AGCGGGCCAGCGCCGAGCGGCTGCCGTACTGCGGTACCAAGCTCGGAGGACCCCGCA 21315  
Qy 365 TGCAGAGAGCAATGAGCTTTGTTCATGAGTGTGCCATACCGCGGCTGCGACGCTC 424  
Db 21314 TGCAGAGAGGCAAGGTGCGGTTCGTGCAAGATGTGAAGTCCGTGGGCGCATTCAGCTGC 21255  
Qy 425 ACCCGAGGCGCATTTGCGGTCTGTTGTATTTGGCAATCCCAAGATGGGTGGCGCA 484  
Db 21254 ACAATCAGGCGCGCTGCGCTTACCTGTGTATTTGGCGCATCCGACCAAGGTTGAGTTT 21195  
Qy 485 TGGCTCTGGGGTTTCATCTGGGCAATCTCACTTTTGGGSAACCGGCGCGCAATAGTTT 544  
Db 21194 TCGCTCTGGGGCTTCGCTGGGGCATCTCAACGTCGCGAGCGGGCGCTGATCGGCT 21135  
Qy 545 TCTTGGGCTCTCGCGGTGAGTTAACACTGGGGCATGCGCTTCCAGACGGTGTGCAGC 604  
Db 21134 TTTTGGGACCAAGGTTCTATGAGTTGCTTATGAGGCAACCTTCCATTCGGCTCAAA 21075  
Qy 605 AGGCGAGAAATTTGGTGAAACTGGTGATGATGAAATGTGTGCGCACTGCAATTGC 664  
Db 21074 CCGCGAAGATCTAGGGGCGGCAATGGGATCATGACGCGCTGTGCACTGACCGGCTAC 21015  
Qy 665 GTGCAAGCGGTGCAAAAACCTCAAGTTATTCAGCGGTAGAGGCAACGATCGTTT 724  
Db 21014 GACCATGCTGATGCTGTGAGCGTGTGATGAGCGTCCGAAACGCTTCCGGCAC 20955  
Qy 725 CTGCAACAATCTGCGGTGCA---CTTCCGGTATGAGGCGGATGCGCTTCTCGTG 781  
Db 20954 GCGAGACCGCGCGCGCTGACCGATGTCGACGTCGAGCTGCGGTGCGATCGCGC 20895  
Qy 782 ACCCGAGAGGCTTGAATTCGGGAGATTATGAAACGTTGGGGCAGACGTCGCAAGC 841  
Db 20894 GCGCGGACCGGCGGCGGTCAAGCTACTGAGACACGCGCGCACGACCGGGTGTGT 20835  
Qy 842 TTTCTGTGCGGTGCTGCGGATTCAGCGCGGTGCGCTGCGCGCATCG 901  
Db 20834 TGTGAGGAACGATCAAGGCGGAGCG---GCGACACGCTGCGCTGGCCGCTTGG 20778  
Qy 902 GGGGCGGCGCGGTGCTGATTTGGGCGAG-----GATGCGCGCTTCAGCG 946  
Db 20777 GCGGCAACCCACGCGGTGCTCTCGGCGAGCAAAAGGCGAGTAGCGCGGCGGAGACACTG 20718  
Qy 947 TTGGGCGCAGAGCTGCGTTTGGCGCTGCGGCAATTCGCTGCGCGCGAGCTAAAC 1006  
Db 20717 TCGGCGCGCGCTGCGTTAGCGAGGCCGACGCGGAGTGGCGCTCGCGCGAGCTGTGCC 20658  
Qy 1007 TGGCATGCTGTCATCATGACACCTTCGCGCGCGAAATTCGCGAGCGCGCTGAGAGC 1066  
Db 20657 TGGCGCTGTGTGTCATTTAGCGCGGCGCGGACCGCGTGTGCGCGGAGCGAAGG 20598  
Qy 1067 TCGCATGCGAAGCTGCAATTTGGCGGCACTTGTCCAGCTTATGACGCTCCCTCCCA 1126  
Db 20597 GCGGCGTGGCGGCAATTCGCGCATTCGTCGAGCTGCTCACTGAGTACCCCGA 20538  
Qy 1127 CCGTTTCGATTAATTTGATGAGGCGTTGGCGGTGCGCGCTGAGCAATGCGCGCG 1186  
Db 20537 CCGTCTCATCTGCTGGGCGGAGGCGGCGGCGCGCGCTGCGGATTTGCGCGCG 20478  
Qy 1187 ATCTGTCTACGCGGCGGCAAAACGCGTGTGCTGCGCATTTGCAAGAGGCGCTCG 1246  
Db 20477 ACCGGGTGCTGCGCTGCACTGCAAGGCTGCGCTGCGCTTGGCTCCGAGAGGCGAGCG 20418

Qy 1247 CCATCTCTTCGCGGACCAACCAACGCGCGGAAATCATAGAGGAGCAAGGCGTGACAG 1306  
Db 20417 CGATGCTTTCGAGACACTGCTCATGCGCGGAACTGCTGCCCAAGGCAATCCGT 20358  
Qy 1307 CGCAGCACTTTTAAAGCCAGGCGTTATGACGCGATGTCGCGCGAAACGAGCACTTTG 1366  
Db 20357 CGGCGGACTTACTGAAGTCGGGGATTTGTGACACCATGCTGCGGAGTATCCCGAGCGCG 20298  
Qy 1367 TTGAAGA 1373  
Db 20297 CAGACGA 20291

Search completed: November 13, 2003, 00:10:39  
Job time : 5538.03 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 15:05:16 ; Search time 407.983 Seconds  
(without alignments)  
9746.164 Million cell updates/sec

Title: US-10-024-370-2

Perfect score: 1473

Sequence: 1 gtggaagaagcgtttccgcac.....gatttcgagcgtttacgcacg 1473

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues 5105512

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: N.GeneSeq\_190un03.\*  
2: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT.\*  
3: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT.\*  
4: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT.\*  
5: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT.\*  
6: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT.\*  
7: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT.\*  
8: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT.\*  
9: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT.\*  
10: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT.\*  
11: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT.\*  
12: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT.\*  
13: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT.\*  
14: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT.\*  
15: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT.\*  
16: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT.\*  
17: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT.\*  
18: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT.\*  
19: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT.\*  
20: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT.\*  
21: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT.\*  
22: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT.\*  
23: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA2001.DAT.\*  
24: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT.\*  
25: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1473	100.0	1473	22	AAH65894
2	1473	100.0	1473	22	AAH65894
3	1473	100.0	1473	22	AAH65894
4	1473	100.0	1473	22	AAH65894
5	1473	100.0	1473	22	AAH65894
6	1473	100.0	1473	22	AAH65894
7	1473	100.0	1473	22	AAH65894
8	1473	100.0	1473	22	AAH65894

9	122.8	8.3	873	23	AAH54163	Pseudomonas aerugi
10	122.2	8.3	876	25	AAH39000	N. gonorrhoeae nuc
11	121	8.2	867	23	AAH53191	Enterococcus faec
12	121	8.2	867	23	AAH53191	Enterococcus faec
13	121	8.2	867	23	AAH53191	Enterococcus faec
14	119.2	8.1	676	24	AAH65750	Helicobacter pylor
15	119.2	8.1	870	23	AAH53651	Helicobacter pylor
16	119.2	8.1	877	24	AAH65721	Helicobacter pylor
17	119.2	8.1	1053	19	AAH14153	H. pylori GHP0 451
18	118.6	8.1	34980	21	AAH21608	Neisseria meningit
19	118.6	8.1	143766	21	AAH14190	N. meningitidis B
20	108.6	7.4	32673	24	AAH27344	Human R128 gene c
21	108	7.3	993	23	AAH56217	Salmonella typhi D
22	102.8	7.0	915	23	AAH52496	E. coli DNA for ce
23	99.8	6.8	3318	23	AAH88487	DNA encoding novel
24	99.8	6.8	3318	23	AAH88487	DNA encoding novel
25	97.8	6.6	867	23	AAH55883	Streptococcus pneu
26	96.8	6.6	654	24	AAH78240	Bacillus clausii g
27	96.2	6.5	864	25	AAH06060	S. pneumoniae type
28	96.2	6.5	867	22	AAH90763	CPE 67 coding sequ
29	96.2	6.5	891	23	AAH53478	Haemophilus influe
30	96.2	6.5	1595	19	AAH296348	S. pneumoniae deri
31	96.2	6.5	19702	19	AAH52140	Streptococcus pneu
32	96.2	6.5	1830121	17	AAH42063	Haemophilus influe
33	96.2	6.5	2162598	25	AAH56454	Streptococcus pneu
34	94.2	6.4	873	24	AAH68650	Streptococcus pneu
35	94.2	6.4	2155561	24	AAH71527	Streptococcus poly
36	93.6	6.4	495269	24	AAH67195	Listeria innocua c
37	93.6	6.4	3011208	24	AAH069245	Listeria innocua c
38	93	6.3	867	23	AAH55630	2CPE 67 coding seq
39	93	6.3	1001	22	AAH90875	Haemophilus influe
40	92.8	6.3	1001	22	AAH90875	DNA sequence upstr
41	92.8	6.3	1001	24	AAH37820	Genomic sequence o
42	91	6.2	2365589	24	AAH30521	Listeria monocytog
43	87.4	5.9	954	18	AAH90220	Magnetospirillum s
44	87.4	5.9	28626	22	AAH28528	Genomic fragment #

## ALIGNMENTS

RESULT 1	AAH65894	standard; DNA, 1473 BP.
ID	AAH65894	
AC	AAH65894	
XX		
XX		
DT	26-SEP-2001	(first entry)
XX		
DE	C glutamicum coding sequence fragment SEQ ID NO: 929.	
XX		
XX	Corynebacterium; amino acid synthesis; vitamin; saccharide;	
KW	organic acid synthesis; ds.	
XX		
OS	Corynebacterium glutamicum.	
XX		
PN	EP1108790-A2.	
XX		
PD	20-JUN-2001.	
XX		
PF	18-DEC-2000; 2000EP-0127688.	
XX		
PR	16-DEC-1999; 99JP-0377484.	
XX		
PR	07-APR-2000; 2000JP-0159162.	
XX		
PR	03-AUG-2000; 2000JP-0280988.	
XX		
PA	(KYOW) KYOWA HAKKO KOGYO KK.	
XX		
XX	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;	
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;	
XX		
XX		
DR	WPI; 2001-376931/40.	



XX 29-NOV-2000.  
PD  
XX  
PF 10-MAY-2000; 2000EP-0109842.  
XX  
PR 27-MAY-1999; 99DE-1024365.  
XX  
PA (DEGS ) DEGUSSA-HUELS AG.  
XX (KERU) FORSCHUNGSZENTRUM JUELICH GMBH.  
PI Tlig Y, Eggeling L, Elkmanns B, Sahn H, Moeckel B;  
XX WPI; 2001-042411/06.  
DR P-PSDB; AAB45789.  
XX  
PT Cloned *Corynebacterium glutamicum* accda gene useful for producing  
PT transformed *coryneform* bacteria producing increased yields of L-amino  
XX acids, especially L-lysine -  
XX  
PS Disclosure; Page 12-15; 20pp; German.  
XX  
XX This invention describes a novel cloned *Corynebacterium glutamicum* DNA  
CC (1), replicable in *coryneform* microorganisms coding for an accda gene.  
CC The invention also describes (1) a polypeptide derived from (1) having  
CC a fully defined 491 aa sequence; (2) *coryneform* microorganisms  
CC transformed with one or more copies of (1); (3) the shuttle vector  
CC p13accda contained in *Corynebacterium glutamicum* DSM 12785; and  
CC (4) a process for producing L-amino acids, comprising culturing a  
CC *coryneform* bacterium that overexpresses the accda gene. *Coryneform*  
CC bacteria transformed with (1) so that they overexpress the accda gene are  
CC useful for producing L-amino acids, especially L-lysine, which are  
CC useful in animal feeds, in human medicine and in the pharmaceutical  
CC industry.

**SQ** Sequence 1473 BP; 275 A; 414 C; 456 G; 328 T; 0 other;

Query Match	100.0%	Score 1473	DB 22	Length 1473
Best Local Similarity	100.0%	Pred. No. 0		
Matches 1473	0	Mismatches	0	Gaps 0

QY	1	GTGGGAGAGCGTTTTC	CCGACTATGGGTGGGGG	CATGAAACACACTTCAGATTTGAGGTC	60
Db	1	GTGGAGAAAGCGTTTTC	CCGACTATGGGTGGGGG	CATGAAACACACTTCAGATTTGAGGTC	60
QY	61	ATGAGCTGGTTTGGAC	CCCTGACAGCTTCATTTCTTGGAAATGAACT	CCCCAATATGAC	120
Db	61	ATGAGCTGGTTTGGAC	CCCTGACAGCTTCATTTCTTGGAAATGAACT	CCCCAATATGAC	120
QY	121	AACCCGATCAAGGCTAT	GCAGACCTTGGAGCGGGGCT	CGAAGCAGAGCCAAATGCGAT	180
Db	121	AACCCGATCAAGGCTAT	GCAGACCTTGGAGCGGGGCT	CGAAGCAGAGCCAAATGCGAT	180
QY	181	GAATCGGTAAATTA	CTGAGAAAGGCAACCGT	GAGGGGATTCGGGTAGCCGTTAATTTGTGC	240
Db	181	GAATCGGTAAATTA	CTGAGAAAGGCAACCGT	GAGGGGATTCGGGTAGCCGTTAATTTGTGC	240
QY	241	GATTTTTCCTTCTCG	GCGGCTCTTTGGGCAAGGTC	CGCTCGGTGGCATCATGAAGCG	300
Db	241	GATTTTTCCTTCTCG	GCGGCTCTTTGGGCAAGGTC	CGCTCGGTGGCATCATGAAGCG	300
QY	301	ATTACCGCGCCACAG	AGCTGAACTCCACATGCTG	GTCTCCCTCCCTTCGCGTGGTGG	360
Db	301	ATTACCGCGCCACAG	AGCTGAACTCCACATGCTG	GTCTCCCTCCCTTCGCGTGGTGG	360
QY	361	CGCATGCAAGAGACAT	CGAGCTTTTGTCA	TGATGATGATGTCATTAACCGCGGCTGTGCAG	420
Db	361	CGCATGCAAGAGACAT	CGAGCTTTTGTCA	TGATGATGATGTCATTAACCGCGGCTGTGCAG	420
QY	421	CGTACACCGAGAGCC	CATTTGCGGTCCGTG	GTGATTTTCCGCAATCCACAGATGGGTGAC	480
Db	421	CGTACACCGAGAGCC	CATTTGCGGTCCGTG	GTGATTTTCCGCAATCCACAGATGGGTGAC	480
QY	481	GCCATGGCCTG	TGGGGTTTCATCTG	GGGCATCTTCTTGGGAAACCGCGGCGGCAGATA	540

[illegible]



XX 11-APR-2001 (first entry)  
DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:25.  
XX  
XX Corynebacterium glutamicum; brevbacterium lactofermentum; MCT;  
KW membrane construction and membrane transport protein; petroleum spill;  
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;  
KW identification; microorganism; fine chemical production; transformation;  
KW genome mapping; genetic engineering; ds.  
OS Corynebacterium glutamicum.  
XX  
XX WO200100805-A2.  
XX  
XX 04-JAN-2001.  
XX  
XX 23-JUN-2000; 2000WO-1B00926.  
XX  
XX 25-JUN-1999; 99US-0141031.  
XX 08-JUL-1999; 99DE-1031454.  
XX 08-JUL-1999; 99DE-1031478.  
XX 08-JUL-1999; 99DE-1031563.  
XX 09-JUL-1999; 99DE-1032122.  
XX 09-JUL-1999; 99DE-1032124.  
XX 09-JUL-1999; 99DE-1032125.  
XX 09-JUL-1999; 99DE-1032128.  
XX 09-JUL-1999; 99DE-1032180.  
XX 09-JUL-1999; 99DE-1032182.  
XX 09-JUL-1999; 99DE-1032190.  
XX 09-JUL-1999; 99DE-1032191.  
XX 09-JUL-1999; 99DE-1032209.  
XX 09-JUL-1999; 99DE-1032212.  
XX 09-JUL-1999; 99DE-1032227.  
XX 09-JUL-1999; 99DE-1032228.  
XX 09-JUL-1999; 99DE-1032229.  
XX 09-JUL-1999; 99DE-1032230.  
XX 14-JUL-1999; 99DE-1032927.  
XX 14-JUL-1999; 99DE-1033005.  
XX 14-JUL-1999; 99DE-1033006.  
XX 27-AUG-1999; 99DE-1040764.  
XX 27-AUG-1999; 99DE-1040765.  
XX 27-AUG-1999; 99DE-1040766.  
XX 27-AUG-1999; 99DE-1040830.  
XX 27-AUG-1999; 99DE-1040831.  
XX 27-AUG-1999; 99DE-1040832.  
XX 27-AUG-1999; 99DE-1040833.  
XX 31-AUG-1999; 99DE-1041378.  
XX 31-AUG-1999; 99DE-1041379.  
XX 31-AUG-1999; 99DE-1041395.  
XX 03-SEP-1999; 99DE-1042077.  
XX 03-SEP-1999; 99DE-1042078.  
XX 03-SEP-1999; 99DE-1042079.  
XX 03-SEP-1999; 99DE-1042088.  
XX  
XX (BADI ) BASF AG.  
XX  
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
XX WPI; 2001-071486/08.  
XX P-PSDB; AAB76522.  
XX  
XX Corynebacterium glutamicum nucleic acids encoding membrane construction  
XX and membrane transport proteins or their portions, useful for typing or  
XX identifying C. glutamicum or related bacteria, and as markers for  
XX transformation -  
XX  
XX Claim 3; Page 173-175; 111pp; English.  
XX  
XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane  
XX construction and membrane transport (MCT) proteins given in AAB76510 to  
XX AAB76847. The MCT nucleic acids and proteins are useful in the  
XX identification of microorganisms which can be used to produce fine

CC chemicals, for modulating fine chemical production in C. glutamicum or  
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or  
CC identification of C. glutamicum or related bacteria, as reference points  
CC for mapping C. glutamicum genome, and as markers for transformation.  
CC AAF68082 and AAF68082 represent sequencing primers which are used in an  
CC example from the present invention.  
XX  
XX SQ Sequence 1575 BP; 309 A; 434 C; 474 G; 358 T; 0 other;  
Query Match 100.0%; Score 1473; DB 22; Length 1575;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTGGAGAAAGCTTTCCGACATGATGTGGGCGATGAAACACACTTCAGCATTAACGCTC 60  
DB 80 GTGGAGAAAGCTTTCCGACATGATGTGGGCGATGAAACACACTTCAGCATTAACGCTC 139  
QY 61 ATGACTGCTGTTTGGAGACCTGACAGCTTCATTTCTTGAATGAATCCCAATATGAC 120  
DB 140 ATGACTGCTGTTTGGAGACCTGACAGCTTCATTTCTTGAATGAATCCCAATATGAC 199  
QY 121 AACCTCAATCAAGCTATGACAGAACCTTGAAGCGGCTCGAAGCAAGCCCAATGAGAT 180  
DB 200 AACCTCAATCAAGCTATGACAGAACCTTGAAGCGGCTCGAAGCAAGCCCAATGAGAT 259  
QY 181 GAATCGGTAATTAATGAGAAAGGACCGTGGAGGGCAATCCGTAAGCGTTATTTGTCC 240  
DB 260 GAATCGGTAATTAATGAGAAAGGACCGTGGAGGGCAATCCGTAAGCGTTATTTGTCC 319  
QY 241 GATTTTCTCTTCTCGGCGGCTCTTCTTGGGACGCTGCGTGGTGCATCATGAAGCG 300  
DB 320 GATTTTCTCTTCTCGGCGGCTCTTCTTGGGACGCTGCGTGGTGCATCATGAAGCG 379  
QY 301 ATTCAACCGCGGCAAGAGCTTAACTCCACTGCTGATCTCCCTCCGCTCCGCGTGGC 360  
DB 380 ATTCAACCGCGGCAAGAGCTTAACTCCACTGCTGATCTCCCTCCGCTCCGCGTGGC 439  
QY 361 CGCATGAGAAAGCAATTCAGCTTTTGTCAATGATGTCATCAACCGCGCTGTGAC 420  
DB 440 CGCATGAGAAAGCAATTCAGCTTTTGTCAATGATGTCATCAACCGCGCTGTGAC 499  
QY 421 CGTACCGGAGGCGCATTTGCGCTTCTGATTTTGGCAATCCCAATGAGTGGC 480  
DB 500 CGTACCGGAGGCGCATTTGCGCTTCTGATTTTGGCAATCCCAATGAGTGGC 559  
QY 481 GCCATGCGCTGCGGCTTCACTGGGCATCTCATCTTTTGGGAAACCGCGCGCACATA 540  
DB 560 GCCATGCGCTGCGGCTTCACTGGGCATCTCATCTTTTGGGAAACCGCGCGCACATA 619  
QY 541 GGTTCCTGCGGCTCTCGGCTGATGAATTAACAATGCGGCTTCCAGACGCTGTG 600  
DB 620 GGTTCCTGCGGCTCTCGGCTGATGAATTAACAATGCGGCTTCCAGACGCTGTG 679  
QY 601 CAGCAGGCGGAAATTTGTAATAAATGCTGTGATTAATGAATTTGTGCGCATCTCAA 660  
DB 680 CAGCAGGCGGAAATTTGTAATAAATGCTGTGATTAATGAATTTGTGCGCATCTCAA 739  
QY 661 TTGGGTGAGCGGTGGCAAAAACCTCAAGTTATTAAGCCGTAAGGCAAGCAATCGT 720  
DB 740 TTGGGTGAGCGGTGGCAAAAACCTCAAGTTATTAAGCCGTAAGGCAAGCAATCGT 799  
QY 721 TTTTCTCAACAACCTCGGCTGAGCACTTCGCTGATGAGGCGGATTCGCTTCGT 780  
DB 800 TTTTCTCAACAACCTCGGCTGAGCACTTCGCTGATGAGGCGGATTCGCTTCGT 859  
QY 781 GACCCGAGAGCGCTGGAATCGGAGATTAATGAAAAGTTGGGGGAGAGAGTGTCAAG 840  
DB 860 GACCCGAGAGCGCTGGAATCGGAGATTAATGAAAAGTTGGGGGAGAGAGTGTCAAG 919  
QY 841 CTTTCTGCTGCGCTGCTGCGCATTAAGCCCGGCTGTGCGCTTGCCTGCGCGCATC 900  
DB 920 CTTTCTGCTGCGCTGCTGCGCATTAAGCCCGGCTGTGCGCTTGCCTGCGCGCATC 979

US-10-024-370-2.rn9

[illegible]

```
Db      1228 TTTTCTCCACAACTCTCGGGGTGGCACTTCCGGTGAAGGAGATTGGCCGCTTCTGCT 1287
Qy      781  GACCCGAGAGGCTCGAATCGGAGAAATTATGAAAAGTTGGGAGGAGAGCTGTCAG 840
Db      1288 GACCCGAGAGGCTCGAATCGGAGAAATTATGAAAAGTTGGGAGGAGAGCTGTCAG 1347
Qy      841  CTTTCTGCTGCGCTGCGGCGGCAATTGAGCCCGGCTGCGGCGGTTGCCCTGCGCGCATC 900
Db      1348 CTTTCTGCTGCGCTGCGGCGGCAATTGAGCCCGGCTGCGGCGGTTGCCCTGCGCGCATC 1407
Qy      901  GGGGGGCGGCGCGGCTGCTGATTTGGGAGAGATGCGCGCTTCAAGCTTGGGCGGAGAG 960
Db      1408 GGGGGGCGGCGCGGCTGCTGATTTGGGAGAGATGCGCGCTTCAAGCTTGGGCGGAGAG 1467
Qy      961  CTGCGTTTGGCGCTGCTGCAATTCGCTGGCGGCGGAGCTAAACCTGCGCATGCTGCC 1020
Db      1468 CTGCGTTTGGCGCTGCTGCAATTCGCTGGCGGCGGAGCTAAACCTGCGCATGCTGCC 1527
Qy      1021 ATCATTCGACACTTCGCGGCGGCAATTGTCGAGGCGGCTGAGAGCTTCGGAGATGGCAAGC 1080
Db      1528 ATCATTCGACACTTCGCGGCGGCAATTGTCGAGGCGGCTGAGAGCTTCGGAGATGGCAAGC 1587
Qy      1081 TCGATTGGCGGCACTTGTCCAAAGCTTATCGACGCTCCCTCCCAACGCTTCCGCTCAT 1140
Db      1588 TCGATTGGCGGCACTTGTCCAAAGCTTATCGACGCTCCCTCCCAACGCTTCCGCTCAT 1647
Qy      1141 ATTGCTCAGGCGCTTGGCGGCTGCGGCGGCTGCGCATGCTCCCGCGCATCTGCTCAGCG 1200
Db      1448 ATTGCTCAGGCGCTTGGCGGCTGCGGCGGCTGCGCATGCTCCCGCGCATCTGCTCAGCG 1707
Qy      1201 GCCGAAAAGCGGCGCTGCTCCGCAATTGCGACAGAGGCGGCTGGGCGCATCTGCTCCGC 1260
Db      1708 GCCGAAAAGCGGCGCTGCTCCGCAATTGCGACAGAGGCGGCTGGGCGCATCTGCTCCGC 1767
Qy      1261 GACACCAACACAGCGCGCGGAAATCATATAGAGGAGACAAAGGCTGACAGCGCACTTTTA 1320
Db      1768 GACACCAACACAGCGCGCGGAAATCATATAGAGGAGACAAAGGCTGACAGCGCACTTTTA 1827
Qy      1321 AGCCAAAGGCTTATTCAGCGGATCGTCCGCCAAACCGAGCACTTTGTAAGAAATTC 1380
Db      1828 AGCCAAAGGCTTATTCAGCGGATCGTCCGCCAAACCGAGCACTTTGTAAGAAATTC 1887
Qy      1381 GGCACAATCAGCAAGCCCTCTCCGATTTGATACATTCGGAGAGGCGGAGACGGAGC 1440
Db      1888 GGCACAATCAGCAAGCCCTCTCCGATTTGATACATTCGGAGAGGCGGAGACGGAGC 1947
Qy      1441 AGTCGCTTCAACGATTTGAGCGTTAGCGAG 1473
Db      1948 AGTCGCTTCAACGATTTGAGCGTTAGCGAG 1980

RESULT 5
AAH68526/c
ID      AAH68526 standard; DNA; 349980 BP.
AC      AAH68526;
XX      XX
XX      26-SEP-2001 (first entry)
XX      XX
XX      C glutamicum coding sequence fragment SEQ ID NO: 7061.
XX      XX
XX      Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX      XX
XX      Corynebacterium glutamicum.
XX      XX
XX      EPI108790-A2.
XX      XX
XX      20-JUN-2001.
XX      XX
XX      18-DEC-2000; 2000EP-0127688.
XX      XX
```

```
PR      16-DEC-1999; 99JP-0377484.
PR      07-APR-2000; 2000JP-0159162.
PR      03-AUG-2000; 2000JP-0280988.
XX      XX
XX      (KYOW ) KYOWA HAKKO KOGYO KK.
XX      XX
PI      Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI      Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX      WPI: 2001-376931/40.
XX      XX
PT      Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT      mutation point of a gene, measuring expression of a gene, analysing
PT      expression profile or pattern of a gene and identifying homologous gene
PT      expression profile or pattern of a gene and identifying homologous gene
XX      XX
PS      Disclosure; SEQ ID NO: 7061; 246bp + Sequence Listing; English.
XX      XX
CC      The present invention provides a number of nucleotide and protein
CC      sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These
CC      are useful for identifying the mutation point of a gene derived from a
CC      mutant of coryneform bacterium, measuring expression amount and
CC      analysing the expression profile or expression pattern of a gene derived
CC      from Corynebacterium bacterium, and identifying a homologue of a gene derived
CC      from coryneform bacterium. Corynebacterium bacteria are useful for producing
CC      amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC      particularly L-lysine. The present sequence is a nucleic acid described
CC      in the exemplification of the invention.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from the
CC      European Patent Office.
XX      XX
SQ      Sequence 349980 BP; 79274 A; 90638 C; 98727 G; 81341 T; 0 other;

Query Match      100.0%; Score 1473; DB 22; Length 349980;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  GTGAGAAAGGTTTCCGATATGATGTTGGGAGATGGAACAACCTTGAGATTGAGCTC 60
Db      281114 GTGAGAAAGGTTTCCGATATGATGTTGGGAGATGGAACAACCTTGAGATTGAGCTC 281055
Qy      61  ATAGACTCGGTTTGGACCTGACAGCTTCAATTTCTTGAATGAATCCCATATGAC 120
Db      281054 ATAGACTCGGTTTGGACCTGACAGCTTCAATTTCTTGAATGAATCCCATATGAC 280995
Qy      121  AACCTCAATCAAGGCTATGAGAGACCTTGAGCGGCTCGAAGCAAGGCCAAATGCGAT 180
Db      280994 AACCTCAATCAAGGCTATGAGAGACCTTGAGCGGCTCGAAGCAAGGCCAAATGCGAT 280935
Qy      181  GAATCGGTAATTAATGAGAAAGGACCGTGAAGGCAATTCGGTAGCGGTAATTTGTC 240
Db      280934 GAATCGGTAATTAATGAGAAAGGACCGTGAAGGCAATTCGGTAGCGGTAATTTGTC 280875
Qy      241  GATTTTCTTCTCTCGCGGCTTCTTGGGACAGCGTCCGCTGAGTCCGATCATGAAGGCG 300
Db      280874 GATTTTCTTCTCTCGCGGCTTCTTGGGACAGCGTCCGCTGAGTCCGATCATGAAGGCG 280815
Qy      301  ATTCAACGCGCACAGAGCTGAATCTCCACTGCTGATCTTCCCTGCTTCGGGTGCG 360
Db      280814 ATTCAACGCGCACAGAGCTGAATCTCCACTGCTGATCTTCCCTGCTTCGGGTGCG 280755
Qy      361  CGCATGACAGAAAGCATTCAGCTTTTGTATGATATGATGATTCCTCAATACCGGCGTGCAG 420
Db      280754 CGCATGACAGAAAGCATTCAGCTTTTGTATGATATGATGATTCCTCAATACCGGCGTGCAG 280695
Qy      421  CGTACCGGAGGAGGCAATTTGCGCTTCTGATGATTTGGCAATCCACGATGGTGGC 480
Db      280694 CGTACCGGAGGAGGCAATTTGCGCTTCTGATGATTTGGCAATCCACGATGGTGGC 280635
Qy      481  GCCATGGGCTTGAGGGGTTCAATCTGAGCATCTCACTTTTGGGAACCGGCGGCAAGTA 540
Db      280634 GCCATGGGCTTGAGGGGTTCAATCTGAGCATCTCACTTTTGGGAACCGGCGGCAAGTA 280575
```

```

QY 541 GGTTCCTGCGTCTGCGTGTGAGTTAAACCACTGGGCAATGCGCTTCCAGCGGTGTC 600
DB 280574 GGTTCCTGCGTCTGCGTGTGAGTTAAACCACTGGGCAATGCGCTTCCAGCGGTGTC 280515
QY 601 CAGCAGCGCGAGAAATTTGGTGAATACTGGTGTGATTTGATGAATTTGTGTGCCATCCAA 660
DB 280514 CAGCAGCGCGAGAAATTTGGTGAATACTGGTGTGATTTGATGAATTTGTGTGCCATCCAA 280455
QY 661 TTGCGTGACGCGGTGGCAAAAACCTCAAGTTATTCACCGGTGATAGGCAACGATTCGT 720
DB 280454 TTGCGTGACGCGGTGGCAAAAACCTCAAGTTATTCACCGGTGATAGGCAACGATTCGT 280395
QY 721 TTTTCTCCAAACAATCTGCGGTGAGCACTTCGCGTATGAGGAGCAATTCGCGCTTCGT 780
DB 280394 TTTTCTCCAAACAATCTGCGGTGAGCACTTCGCGTATGAGGAGCAATTCGCGCTTCGT 280335
QY 781 GACCCGACAGAGCCTTGAAATCGGGAGATTTATGAAAAGTTGGGGGACAGACGTCTCAAG 840
DB 280334 GACCCGACAGAGCCTTGAAATCGGGAGATTTATGAAAAGTTGGGGGACAGACGTCTCAAG 280275
QY 841 CTTTCTGCGTGGCGGTGCGGCTGAGCAATTCGCGGTGAGCGGTGCGGTGCGGTGCGGTATC 900
DB 280274 CTTTCTGCGTGGCGGTGCGGCTGAGCAATTCGCGGTGAGCGGTGCGGTGCGGTGCGGTATC 280215
QY 901 GGGGGCCGCGCCGCTGCGTGTGATTTGGGACAGATTCGCGCTTCAAGCTTGGGCGGACAGAG 960
DB 280214 GGGGGCCGCGCCGCTGCGTGTGATTTGGGACAGATTCGCGCTTCAAGCTTGGGCGGACAGAG 280155
QY 961 CTGCGTTTTGCGCGTGTGAGCAATTCGCGTGGCGCGGAGCTAAACCTGCGCATTCGTGTC 1020
DB 280154 CTGCGTTTTGCGCGTGTGAGCAATTCGCGTGGCGCGGAGCTAAACCTGCGCATTCGTGTC 280095
QY 1021 ATCATTCAGACCTTCGCGCGCCGCAATTTGTGCGAGCGCGCTGAGAGAGTCCGCAATCCGACG 1080
DB 280094 ATCATTCAGACCTTCGCGCGCCGCAATTTGTGCGAGCGCGCTGAGAGAGTCCGCAATCCGACG 280035
QY 1081 TCGATTGCGCGACCTTGTGCAAGCTTATCGACGCTCCCTCCCAACGCTTTCGGTCAAT 1140
DB 280034 TCGATTGCGCGACCTTGTGCAAGCTTATCGACGCTCCCTCCCAACGCTTTCGGTCAAT 279975
QY 1141 ATTGCTCAGAGGCGTTGGCGGTGCGCGCTGCGCATTCGCGCGCATTCGCTTCAACGCG 1200
DB 279974 ATTGCTCAGAGGCGTTGGCGGTGCGCGCTGCGCATTCGCGCGCATTCGCTTCAACGCG 279915
QY 1201 GCCGAAAACGCGTGTGCTGCGCATTCGCGCATTCGCGCATTCGCTTTCGCGC 1260
DB 279914 GCCGAAAACGCGTGTGCTGCGCATTCGCGCATTCGCGCATTCGCTTTCGCGC 279855
QY 1261 GACACCAACCAACGCGCGGAAATTCATAGAGCGAACAAGGCGTGCAGGCGGACGACTTTTA 1320
DB 279854 GACACCAACCAACGCGCGGAAATTCATAGAGCGAACAAGGCGTGCAGGCGGACGACTTTTA 279795
QY 1321 AGCCAAAGGCGTTATCGACGCGATTCGCGCAACCGAGCACTTTTGTGAAGAAATTCCTC 1380
DB 279794 AGCCAAAGGCGTTATCGACGCGATTCGCGCAACCGAGCACTTTTGTGAAGAAATTCCTC 279735
QY 1381 GGCACAAATCAGCAAGCGCTTCTCGCAATTTGATTAACAATCCGAGAGAGGCGGAGCGGAC 1440
DB 279734 GGCACAAATCAGCAAGCGCTTCTCGCAATTTGATTAACAATCCGAGAGAGGCGGAGCGGAC 279675
QY 1441 AGTCGCTTCAACGATTTGAGGCTTTAGGCGAG 1473
DB 279674 AGTCGCTTCAACGATTTGAGGCTTTAGGCGAG 279642

```

```

RESULT 6
AAH51981
ID AAH51981 standard; DNA; 1488 BP.
AC AAH51981;
XX
DT 04-SEP-2001 (first entry)

```

```

XX XX
DE Mycobacterium tuberculosis potential drug target gene SEQ ID 35.
XX XX
KW Drug target; growth; organism viability; characterisation; dr.
XX XX
OS Mycobacterium tuberculosis.
XX XX
PN WO200135317-A1.
XX XX
PD 17-MAY-2001.
XX XX
PF 13-NOV-2000; 2000WO-US31152.
XX XX
PR 12-NOV-1999; 99US-0165086.
XX XX
PR 12-NOV-1999; 99US-0165124.
XX XX
PR 01-FEB-2000; 2000US-0179531.
XX XX
XX (REGC ) UNIV CALIFORNIA.
XX PA
XX PI Eisenberg D, Roststein SH, Marcotte EM;
XX PI WPI, 2001-329193/34.
XX DR P-PSDB; AAG81130.
XX PT
XX PT Identifying nucleotide or polypeptide sequence for use as drug target,
XX PT involves providing algorithm that analyzes a functional relationship
XX PT between nucleotide or polypeptide sequences, and comparing the
XX PT sequences.
XX PS
XX PS Disclosure; Page 74; 207tp; English.
XX XX
CC This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analyzing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterizing the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.
XX XX
SQ Sequence 1488 BP; 220 A; 509 C; 524 G; 235 T; 0 other;
XX XX
Query Match 21.2%; Score 312.6; DB 22; Length 1488;
Best Local Similarity 54.0%; Pred. No. 1.5e-78;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

```

```

Db      332  TCGAAGAGGACGCGCTTCTTCTGAGATGATGAGATGCTGCGGCATCCAGCTGC 391
Qy      425  ACCGCGAGGCGCATTTTCCGTTCTGATTTTTCGCAATTCACAGATGGGTGGCCCA 484
Db      392  ACAACGAGGCGCGCTGCTTACCTGCTATTTTGGCCATCCAGCACGCGGTGAGTTT 451
Qy      485  TGGCTCTGTTGGGTTCACTGTCGATCTTCTTGGGAAACCGCGCGCGAGATAGTTT 544
Db      452  TCGGCTGTTGGGCTCGCTGGGCACTTCACGTCGCGAGCCGCGCCCTGATCGCT 511
Qy      545  TCTGGGCTCTGCGCTGTTGAGTTAAACCACTGGGCGATCGCTTCCAGAGGTGTGACG 604
Db      512  TTCTGGAGCACAGGCTTATATGATTTGCTATGAGGACCCCTTCCATCCGCGCTCAA 571
Qy      605  AGGCGAGATTTGGTGAATCTGTTGATTTGATGATGATTTGTTGCTGCTCCATTTGC 664
Db      572  CCGCGAGATCTACGCGCGCATGGGATCATGACGCGGTGCTGCTGACATCGGCTTAC 631
Qy      665  GTGACGCGGTGGCAAAACCTTCAAGTTTATTCAGCCGTTAGAGGCAAGATGCTTTT 724
Db      632  GACCGATCTGATCTGCTGCTGAGCGTCTATGACGCTCCGACCTTCCGAC 691
Qy      725  CTCCAGCAACTCTGCGCGTGGCA---CTTCCGTTGATGAGGCGCATTCGCGTTCTGCTG 781
Db      692  CGCAGACCCCGCGCCCTTACCCGATGTCACAGTGGGACTCGGTGGTGGATCGGCC 751
Qy      782  ACCGCGAGAGGCTTGGAAATCGGGAGATTTAGAAACCTTGGGGGAGAGCTGCTGAAC 841
Db      752  GCGCGAGACCGGCGCGGCTCAGGACAGTACTGCGACAGCGGCCCAACCGGCGTGTGT 811
Qy      842  TTTCTGTTGCGCGCTGCTGCTGCTGATGAGCCGCGCTGCTGCGGTTGCGTGGCGCATCG 901
Db      812  TGTAGGAACCGATTAAGAGCGAAGCG---GCGACACGCTGCTGCGCTGGCCGCTTTTG 868
Qy      902  GGGGCGCGCCCGTGTGCTGATTTGGGCGAG-----GATCGCGCTTTCAGCGC 946
Db      869  GCGGCGAACCCAGCGTGTGCTTCCGCGCGAGAAAGGCGAGTAGGCGCGCGGGAGAACATCG 928
Qy      947  TTGGGCGCAGAGAGCTGCGTTTGGCGCTGCTGCGCATTTTCTGTCGCGCGCGAGCTAAAC 1006
Db      929  TCGGGCGCGCTGCTGCTGCTGAGCGAGCGCGAGCGGCGATGCGCTGCGCGCGAGCTGTGCC 988
Qy      1007  TGGCGATGTTGCTGATTCAGACACCTCGCGCGCGAAATTTGTCGAGCGCGGTGAGAGC 1066
Db      989  TGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1048
Qy      1067  TCGGATCGCAAGCTGCAATTTGGCGCGACCTTGTTCAGAGTTTTCAGAGCTTCCCTCCCA 1126
Db      1049  GCGGCGCTGCGCGCGCGAGATTCGCGCATTTGCTGCGAGCTGCTGCAAGCTGATACCCCGA 1108
Qy      1127  CCGTTTCGCTATTTATTTGTCAGGGCGTTTGGCGGTGGCGCGCTGCGCATGCTGCGCGCG 1186
Db      1109  CCGTTCGATCTGCTGCGCGCGCGAGCGCGCGCGCGCGCGCGCTGCGATGTTGCGCGCGCG 1168
Qy      1187  ATCTGCTGACGCGCGCGCGAAACGCGTGGCTGTTCGCAATTTGTCACAGAGGCGCTCGG 1246
Db      1169  ACCGGGTGCTGCGCGCATTCACGCGGTGCTGCGCGCTTGTCTTCCAGAGAGCGACGCG 1228
Qy      1247  CCATCTCTTCCCGCGACACCAACCAACCGCGCGGAAATCATAGAGCGACAGGCGGTGACG 1306
Db      1229  CGATGTTGTTCCGAGACACTGCTCATGCGCGCGAACTGCTCGCGCGCAAGGCGATCCGGT 1288
Qy      1307  CGCAGCGACTTTTAAAGCAAGGCTTATGACGGGATGCTGCGCGAAACCGAGCACTTTG 1366
Db      1289  CGGCGGACTTCTGAATGCGGGGATTTGTCAGACCATGCTGCGGAGTACCCCGAGCGCG 1348
Qy      1367  TTGAGA 1373
Db      1349  CAGACGA 1355

```

RESULT 7  
AA199683/c

```

ID      AA199683 standard; DNA: 4403765 BP.
XX
AC      AA199683;
XX
DT      15-JAN-2002 (first entry)
XX
DE      Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KM      Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
XX      variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS      Mycobacterium tuberculosis..
XX
PN      US6294328-B1.
XX
PD      25-SEP-2001.
XX
PF      24-JUN-1998; 98US-0103840.
XX
PR      24-JUN-1998; 98US-0103840.
XX
PA      (GENO-) INST GENOMIC RES.
XX
PI      Fleischmann RD, White OR, Fraser CM, Venter JC;
XX      WPI: 2001-647261/74.
XX
PT      Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT      determining the nucleotide sequence of the strain at positions in the
PT      genome corresponding to positions where M. tuberculosis strains CDC
PT      1551 and H37Rv differ.
XX
PS      Claim 4; SEQ ID NO 2; 3bp + Sequence Listing; English.
XX
CC      The invention relates to evaluating strain variation within and between
CC      different populations of the tuberculosis bacterial pathogen,
CC      Mycobacterium tuberculosis or related Mycobacterium by determining the
CC      nucleotide sequence of the first strain at positions in the complete
CC      sequence of the genome that correspond to positions that differ in the
CC      nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC      H37Rv (AA199682). The method is useful for evaluating strain variation of
CC      M. tuberculosis and has valuable application in the fields of
CC      tuberculosis genetics, epidemiology, patient treatment and epidemic
CC      monitoring.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from USPTO
CC      at seqdata.uspto.gov/sequence.html?docID=6294328B1.
XX
SQ      Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match      21.2%; Score 312.6; DB 22; Length 4403765;
Best Local Similarity 54.0%; Pred. No. 5.4e-77; Indels 21; Gaps 3;
Matches 717; Conservative 0; Mismatches 589;

Qy      65  ACTGCTTTTGAACCTGACAGCTTCAATTTCTTGAATGAATCTCCCAATAGACAAC 124
Db      1008148  ACGGCTCTAGAACCGGGGATCTTCTGCACTGAGATAGAGACCGCTGCGGTGCGG 1008089
Qy      125  TCAATCAAGGCTATGACAGACCTTTGAGCGGCTTCAGAGCAAGCCAAATGCATGAAT 184
Db      1008088  TAGCGCACTCTTATGCGGAGAGTGGCCGCGCTCGGCGCGCACCGCGCGGACGAT 1008029
Qy      185  CGGTAATTACTGGAAGAGCGACGTTGAGAGGCAATTCGCGTATTTTGTCCGATT 244
Db      1008028  CGGTGACAGCCGTAAGAGAGCGGTAATTCGGGCGCGGAGTGGCTGTGCTGTGAGT 1007969
Qy      245  TTCTCTCTGCGGCTTCTTTGGGACAGGTGCGGTGCTGCGATCATGAAGCGATTC 304
Db      1007968  TCGACTTCTGCGGCGCTGCTGATTTGGGAGGAGGCGCGGAGGATCAACCGCGCGTGC 1007909
Qy      305  ACCGCGCACAGAGCTGAATCCCATGCTGTGTTCTCCCTGCTTCCGTTGCTGCGGCA 364
Db      1007908  AGCGGCGACCGCGAGCGGCTGCGCTACTAGCGGTCAACCAAGCTCGGAGGACCCCGCA 1007849

```

QY	365	TTCCAGAAAGACATAGAGCTTTTTCATGATGCTGCATTAACCGCGGCTGTGACGCTC	424
Db	1007848	TTCCAGAAAGGCAACGGTCCGCTTTTCGACAGATGCTGATAGATCGCTGCGCATTCAGCTGC	1007789
QY	425	ACCGGAGGCGCATTTGCGCTTCTGGGTGATTTGGGAATTCGACATCCGATGGGTGCGCA	484
Db	1007788	ACAACACAGCGCGCTGCGCTTACCTGCTGCTATTTTTCGCGATTCGACACAGGCTGAGTTT	1007729
QY	485	TGGCTCTGTGGGGTTCATCTGGGACCTTCACCTTTTTCGGAACCGCGCGAGATAGGTT	544
Db	1007728	TGCGCTCGTGGGGCTCGTGGGGGATCTCACCGTGGCCGACCGGGCGCCTGATCGGCT	1007669
QY	545	TCCTGGGCTCTCGCGTGTGAGATTAAACCACTGGCATGCGCTTTCAGACGCTGTGACG	604
Db	1007668	TTCTTGGGACCAACGGGCTCTATGAGTTGCTCTATGCGGACCCCTTCCATCCGGCTCCAAA	1007609
QY	605	AGCGGGAATTTTGGTGAATACTGGTGTGATTTGATGAAATTGTGTGCACTCCAAATTGC	664
Db	1007608	CCGCGGAATTTCTACGGCGGACATGGAATTCAGACGGGTGTTGCACTTGACACGGCTAC	1007549
QY	665	GTGCAACCGGTGGCAAAAACCTCAAGTTATTACCGCGGTGAGGCAACGATCGTTTTT	724
Db	1007548	GACCATCTGTGATGTGCTGCTGTGACGGTGTCTATGACGCTCCGAAACGCTTCCGGCAC	1007489
QY	725	CTCCAAACAATCTTGGCGGTGCA---CTTCGGGTGATGAGAGCGCATTTGCGGTTCTGCTG	781
Db	1007488	CGCAGACCGCCGCGCCGTACCCGATGCGCAATGCGCAATGCGGATCTCGGTGGCATGCGCC	1007429
QY	782	ACCGGCAAGGCTGTGAATCGGGAGATTATGAAAACCTTGGGGCAGACGTCTCAAC	841
Db	1007428	GCGCGGACCGGCGCGGCGCTCAGGCAAGCTACTGCAACGCGGACACGACCGGCTGTTGT	1007369
QY	842	TTTCTGAGCGCGTGTGCGGCATTGAGCCGGGCTGTGCGGTTGCCCTGCGCGGCAATCG	901
Db	1007368	TGTCAAGAACGATTCAGGCGAAGCG---GCAACACGCTGCTGACGCTGCGCTTTG	1007312
QY	902	GCGGCGCGGCGCGGTGTGCTGATTGGGCAAG-----GATCGCGCTTCAACG	946
Db	1007311	GCGGCAACCCACGGTGTCTCGGCGAGCAAGAGGACGATGAGGGCGGGGAAGCACTG	1007252
QY	947	TTGGGCGCGAGAGCTGCTTTTGGCGGCTGTGCGATTTGCTGTGGCGCGAGCTAAAC	1006
Db	1007251	TGCGGCGCGCTGTGCTTACGCAACCCGACGCGGGAATGCGCTGCGCGAGCTGTGCC	1007192
QY	1007	TGCGGATGTGTCCATATTCACACATCTCGCGCGCGGATTTGTGCGCAGCGGCTGAGAC	1006
Db	1007191	TGCGGCTGTGTCTGTATTCATCCCGCGCGGACCCGGTGTCTGCGCGCACCGCAACAG	1007132
QY	1067	TCGGCATCGCAAGCTCGATTGCGGCACTTGTCCAGGCTTATGAGAGCTCCCTCCCA	1126
Db	1007131	GCGGCGTGGCGCGGACAGTCCGCGATTGCTTGGCGAGCTGTGATACCTGGAATCCCGA	1007072
QY	1127	CCGTTTGGCTATTATTGTTCAGGCGGCTGTGCGGCTGGCGGCTGCGCATGCTGCCCGC	1186
Db	1007071	CCGTTGTGATCTCTGTGGCGCAAGGCAAGCGCGCGGCGGCTGGGATGTTGGCGCGCG	1007012
QY	1187	ATCTGTGTACGCGCGCGAAACCGCGTGTCTGCAATTGCGCACAGAGGGCGCTCGG	1246
Db	1007011	ACCGGCTGTGGCGCACTCCACGCTGTGCGCGCTTGTGCTCCGAGAGGACGACG	1006952
QY	1247	CCATCTCTTTCGCGGACCAACACACGCGCGGAAATCATAGAGGCAAGAGCGGTGACG	1306
Db	1006951	CGATCTGTTCGAGACACTGTCTATGTGCGCGCAACTGTGCGCGCCCAAGGATTCGGT	1006892
QY	1307	CGACACGACTTTTAAGCCAAAGGCTTATTCAGAGGATGCTGCGCGAAACGAGACATTTG	1366
Db	1006891	CGGCGGACCTTACTGAAGTGGGAGATTGTTCAGACACATCTGTGCGGGAATTACCCGACCGCG	1006832
QY	1367	TTGAAGA 1373	
Db	1006831	CAGACGA 1006825	

```

RESULT 8
ID AA19682/c
XX AA19682 standard; DNA; 4411529 BP.
AC AAI9682;
XX
XX
DT 15-JAN-2002 (first entry)
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KM variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
PN US6294328-BL.
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
PR 24-JUN-1998; 98US-0103840.
PA (GENO-) INST GENOMIC RES.
PI Fleischmann RD, White OR, Fraser CW, Venter JC;
DR WPI; 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ -
PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI9683) and
CC H37Rv (AAI9682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?poCID=6294328B1.
CC XX
SQ Sequence 4411529 BP; 75855 A; 1449983 C; 1444602 G; 758379 T; 0 other;
DB
QY Query Match 21.2%; Score 312.6; DB 22; Length 4411529;
Beat Local Similarity 54.0%; Pred. No. 5.4e-77;
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;
DB
QY 65 ACTGGGTTTGTAGACCCTGCAGACTTCATTCTTGGAATGAATACTGCCAATATGACAACC 124
DB 1008147 ACCGGCGGTCAAGACCGGGAATCTTTCGTACGTGGAGAAGCACCCTGGCGGTGCGG
    125 TCATTAACAGGCTATGCAGAGAACCTTAGAGGGGCGCTGAAGCAAAGGCCAANTGSGATGAAT 184
DB 1008087 TAGCCGACTCTTAAGCCCCGGAGAGCTGGCCCCTGCGGGCGCACCGCGCGGACGAAT 1008028
QY 185 CGGTAAATTAATGCAGAGGACACCGTGGAGGGGCAATCCGGTAGCCGTAATTTHGCCGATT 244
DB 1008027 CGGTAGAGACCGGTGAAGGAGAGOGSTATTTCGGGGCGGGGGTGGCGGTGTGCGCTGTAGT 1007968
QY 245 TTTCCTTCCTCGGGGGGTTCTTTGGGCAAGCGTCCGCTCGGTGCGCATATGAAGGCGATTC 304
DB 1007967 TGCACTTCCTCGGGGGGCTCGCATTTGGGGTGGCAGGCGCCAAGCATACCGCGCGCTCG 1007908

```



```

QY 305 ACCGCGCAGAGCTGAACTCCCACTGCTGCTCCCTCTCCGCTGGTGGCGCA 364
Db 1007907 AGCGGCGCAGCGCCAGCGGCTCCGCTACTGCTCAACAGCTGGGAGGCAACCGCA 1007848
QY 365 TGCAGAAAGCAATGAGCTTTTGTCAATGATGCTCAATACCGCGGCTGTGACGCTC 424
Db 1007847 TGCAGAAAGCAATGAGCTCCGCTTTCAGAGATGCTCAATACCGCTCCGCTCCAGCTGC 1007788
QY 425 ACCGCGAGCGCATTTTGGCTTCCGCTGATTTTGGCAATCCGCAATGGGTGGCGCA 484
Db 1007787 AACACAGAGCGCGCTCCGCTCACTGCTATTTTGGCCATTCGACACGAGGTGGAGTTT 1007728
QY 485 TGGCTCTGAGGAGTTCATCTGGGCACTCACTTTTGGCGAAACCGCGCGCAGATAGGTT 544
Db 1007727 TGGCTCTGAGGAGTTCGCTGGGCACTCACTGCTGGCGAGCGCGCGCTCCGCTGAGCT 1007668
QY 545 TCTGGGCTCTGCTGCTGAGTTAACACTGGGCAATGCGCTTCAACAGCTGTGACGC 604
Db 1007667 TTTCTGGGCAACAGGCTCTATGAGTTGCTATGCGACCCCTTCCCATCCGCGCTCAAA 1007608
QY 605 AGCGGAGAAATTTGGTGAATACTGGTGTGATGAGAAATTTGTGCGCACTCCCAATTGC 664
Db 1007607 CCGCGAAGATCTACGCGCGCGCATGGGATCATGACGCGCTGCTGCACTGACCGGCTAC 1007548
QY 665 GTGCAAGCGGTGCAAAAAACCTCAAGTTATTCAAGCGGTAGAGGCAACGAGTCGTTTTT 724
Db 1007547 GACCGATGCTGATGCTGCTGTGAGCGGTCTCATGAGAGCTGCCGACCGCTCCGGGAC 1007488
QY 725 CTCGACAACTCTGCGCGTGGCA---CTTCCGCTATGAGAGCGGATTTGCGCTTCTGCT 781
Db 1007487 CGCAACCCCGCGCGCGCGCTGACCGGATGCTCCACGCTGGGAACTCGTGTGGCATGCGCC 1007428
QY 782 ACCCGCAGAGGCTGGAATCCGGGAGATTATGGAACGTTGGGGGCAACGCTCGCAAGC 841
Db 1007427 GCGCGGACCGCGCGCGCGCTGACGCTACTGCAACGCGCGCAACCGAGGCTGTGT 1007368
QY 842 TTTCTGCTGCGCGCTGCTGCGCATTTAGCGCGCTGTGCGCTTCCCTGCGCGCATCG 901
Db 1007367 TGTCAAGAAACGATCAAGCGCGAAGG---GCGACACGCTGCTGGCGCTGCGCGCTTTG 1007311
QY 902 GGGGCGCGCGCTGCTGCTGATTTGGCG---GATGCGCGCTTCAAGC 946
Db 1007310 GCGGCGCAACCCGCTGCTGCTGCGCGCAAGGCGAGTAGCGCGCGGCGGACACTGC 1007251
QY 947 TTTGGCGCGGAGGAGCGCTTTTGGCGCTGCTGCGCATTTGCTGGCGCGCGCAATACC 1006
Db 1007250 TCGGCGCGCGCTGCTTACGCGAAGCCCGAGCGGAGTGGCGCTGCGCGCGCTGTGC 1007191
QY 1007 TCGCGATGCTGCTCATCATCATGACACCTCCGCGCGCGAATTTGCGAGCGCGTGAAGC 1066
Db 1007190 TCGCGCTGCTGCTGCTGCTGCTGCTGCGCGCGCGCGCGCTGCTGCTGCGCGCGAAGC 1007131
QY 1067 TCGCGATGCGAAGCTGATTTGGCGCGCACTTGTCAAGTTATGCAAGCTCCCTCCCA 1126
Db 1007130 GCGGCGCTGCGCGCGCGCAATGCGCGATTTGCTGCGCGAGCTCCCTCAAGCTAATCCCA 1007071
QY 1127 CCGTTTCCGCTCATTTTGGTCAAGGCGCTTTGGCGCGCGCGCGCTGCGCGCATTCGCGCGCG 1186
Db 1007070 CCGTGTCAATCTGCTGCGCGCGAGGCGCGCGCGCGCGCGCGCTGCGAGTGTGCCCGCG 1007011
QY 1187 ATCTGCTACCGCGCGCGCGAAGCGGTGCTTCCGATTTGCGACGCGCGCGCTCGG 1246
Db 1007010 ACCGCGTGTGCGCGCGCGCATTCACAGCGTGGCTGGCGCTTGTGCTCCGAGAGCGCAGG 1006951
QY 1247 CCATCTCTTCCGCGCACACCAACGCGCGGAAATCATAGAGCGACAGGCGTGCAG 1306
Db 1006950 CGATCGTGTCCGAGACACTGCTCATGCGCGCGAATCGCTGCGCGCGCAAGCATCCGAT 1006891
QY 1307 CGCAGCGACTTTAAGCGCAAGGCTTATGACGCGATTCGTCGCCGCAACCGAGCATTTG 1366
Db 1006890 CGGCGGACTTACTGAAGTGGGAGTTGTGACACCATGCTGCGCGAGTACCCCGAGCGCG 1006831
QY 1367 TTGAAGA 1373

```

```

Db 1006830 CAGACGA 1006824

RESULT 9
AAS54163
ID AAS54163 standard; DNA; 873 BP.
XX
AC AAS54163;
XX
DT 13-FEB-2002 (first entry)
XX
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #294.
XX
KW Antisense; ds: prokaryotic cellular proliferation gene;
XX antibiotic; antibacterial; drug design.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
XX
PR 23-MAY-2000; 2000US-206848P.
XX
PR 26-MAY-2000; 2000US-207727P.
XX
PR 23-OCT-2000; 2000US-242578P.
XX
PR 27-NOV-2000; 2000US-253625P.
XX
PR 22-DEC-2000; 2000US-257931P.
XX
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-), ELITRA PHARM INC.
XX
PI Hazelbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI: 2001-611495/70.
XX
DR P-PSDB; AAU36304.
XX
PT New polynucleotides for the identification and development of
PS antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Claim 27; Seq ID No 7800; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence encodes an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 873 BP; 152 A; 285 C; 276 G; 160 T; 0 other;

Query Match 8.3%; Score 122.8; DB 23; Length 873;
Best Local Similarity 52.8%; Pred.No. 1.5e-24;
Matches 265; Conservative 0; Mismatches 237; Indels 0; Gaps 0;
QY 179 ATGAATCGTAATTACTGAGAAAGCAACGTGAGGCGATTCGGTAGCCGTTATTTTGT 238

```



```

Db      326 AGGACGCGTGTATCGCATGAGCGCAAGCTGACGGGCAATGCGGTGTGCGCTTGCCCT 385
Qy      239 CCGATTTTCTTCTCTCTCGGCGGTTCTTTGGGCACGGTCGGTGGCGCATGTAAG 298
Db      386 TCGAGTTCTCTTCTTATGAGGGGTTGCAATGGGCGCATTCGCGGAGCGCTTCTGCGG 445
Qy      299 CGATTCACCGCGGCACAGAGCTGAAACTCCACTGCTGTCTCCCTGCTTCCGGTGTG 358
Db      446 CAGCCAAAGTCGCTTGTAGAGAGCGGTGCGCTGATCTGTCTTCTCCGCTCCGGGCGG 505
Qy      359 CCGCATGACAGAGAACATTCAGCTTTTGTCTATGATGTGTCCATTAACCGCGCTGTGC 418
Db      506 CCGCATGACAGAGAGGCTGATCTCGCTGATGACAGATGACCAAGACTCGCGGCTTCTG 565
Qy      419 AGCGTACCGCGGAGCGCATTTGCGCTGCTGTGTATTTGGGCAATCCACAGATGGTG 478
Db      566 CCGCGCTGCGGCAGAGAGCATCCGCTGTCTCGGTATTTGACCGACCGGCTTACGGCG 625
Qy      479 GCGCCATGCGCTCTGTGGGGTTTCATCTGGGCATCTCACTTTGGGGAACCGGCGGCGAGA 538
Db      626 GCGTTTCCGCGCAGCGCTGGGAGTCTGGGAGCTGATCGTCCGCAACCGAAGCGCTGA 685
Qy      539 TAGGTTTCTGAGTCTCGCGTGTGTGAAGTTAACCATGGGCAATGCGCTTCCAGACGTG 598
Db      686 TCGGCTTCCCGGCTCTCGCGTGTGATGACAGACCGTCCGCGAAGGCTTCCGGAAGGCT 745
Qy      599 TGCAGACGCGGAGAAATTTGTGAATACTGTGTGATTTATGGAATTTGTGTGCCACTTC 658
Db      746 TCCAGGTATGACAGATCTCTCTTGTAGCATGCGCCATGCAATGATGTGATCGTCCG 805
Qy      659 AATTGCGTGACAGCGGTGCAAA 680
Db      806 AGTTGCGGCGCGGCGCTGCGCAA 827

```

RESULT 10  
ABZ39000  
ID ABZ39000 standard; DNA; 876 BP.

```

XX      AC      ABZ39000;
XX      DT      07-MAR-2003 (first entry)
XX      DE      N. gonorrhoeae nucleotide sequence SEQ ID 2589.
XX      KW      Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX      OS      Neisseria gonorrhoeae.
XX      PN      WO200279243-A2.
XX      PD      10-OCT-2002.
XX      PE      12-FEB-2002; 2002WO-IB02069.
XX      PR      12-FEB-2001; 2001GB-0003424.
XX      PA      (CHIR-) CHIRON SPA.
XX      PI      Fontana MR, Pizza M, Massignani V, Monaci E;
XX      DR      MPI; 2003-058415/05.
XX      P-PSDB; ABP78030.
XX      PT      New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX      FT      medicament for treating or preventing N. gonorrhoeae infection -
XX      PS      Disclosure; Page 377; 815pp; English.
XX      CC      The present invention relates to proteins from Neisseria gonorrhoeae.
XX      CC      Also disclosed are the nucleic acid molecules encoding the proteins and
XX      CC      antibodies that specifically bind to the proteins. The composition

```

CC comprising the protein, nucleic acid or antibody is useful for the  
CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
CC infection; this may be in the form of a vaccine or gene therapy.  
CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid  
CC molecules of the invention.

XX SQ Sequence 876 BP; 183 A; 216 C; 272 G; 205 T; 0 other;

Query Match 8.3%; Score 122.2; DB 25; Length 876;  
Best Local Similarity 51.6%; Pred. No. 2.2e-24;  
Matches 280; Conservative 0; Mismatches 263; Indels 0; Gaps 0;

```

Qy      121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTCCAGAGCAAGCCAAATCCAT 180
Db      274 AAAGACAGCAAAATATCCGACCGCTTGTAGTGGGACGACGACTAACGGGAAAGT 333
Qy      181 GAATCGTAATTTACTGAGAAAGGACCCGTGAGAGGCAATTCGGGTACCGTTATTTGTCC 240
Db      334 GACGCGCTGTGTGATGAGAGGGAGTATGAAACGGTCTGCCGCTGTGTTGCCGCTTT 393
Qy      241 GATTTTCTCTCTCTGCGGCTTCTTTGGCACGCTCGCGTGGCCATATGAGCG 300
Db      394 GAATTCGCGCTTATTCGGGCGGTTGATGATGGTTCCGTTGTGGCGAGCCGTTGTAAGGAT 453
Qy      301 AATCACCGCGCACAGAGCTGAAACTCCCACTGCTGTCTCCCTGCTTCCGGTGTGCG 360
Db      454 GTCCGCGCTGCGGTTGCCGATATTTGTCTGTGTGTGTGTGGCGGCTTCCGCGCGCG 513
Qy      361 CCGATGACAGAAACATATGAGCTTTTGTATGATGTGTCCATTAACCGCGCTGTGCGAG 420
Db      514 CGTATGACAGAGGGGTATTAATCGCTGATGACATGAGAAACAGATGCCGCGCTGCAT 573
Qy      421 CGTACCGCGAGGCGCATTTGCCGTTCTGTGTATTTGGCAGATCCACAGATGGTGGC 480
Db      574 TTGCTGACGAAAAACCGCTGCGTTATATCGGTGTGACCGATCCACTATGAGGCGGC 633
Qy      481 GCATGCGCTCGTGGGGTTTCATCTGGGCACTCACTTTTGGGGAACCGGCGCGAGATA 540
Db      634 GTATCTGCACTTCCGCAATTTTATGCGCATTCGTCTTCCGAAACGGAACCGCTGATC 693
Qy      541 GATTTCCTGAGTCTCTCGCGTGTGAGTTAACCATGCGGCAATGCGCTTCCAGACGTTG 600
Db      694 GATTTCGCGGTCGCGCGGTGATGAGCAGACGGTGGCGCAAGGCTCCGGAAGGCTTC 753
Qy      601 CAGCAGCGGAGGAATTTGGTGAATAAATGCTGTGATTTGATGGAATTTGTCTCCACTCA 660
Db      754 CAGCGCGCGAGTTTCTGTGGAATAAAGCGCGATGACACGATTTGTGACCGCGCGAT 813
Qy      661 TTG 663
Db      814 ATG 816

```

RESULT 11  
AAS53191  
ID AAS53191 standard; DNA; 867 BP.  
XX AC AAS53191;  
XX DT 13-FEB-2002 (first entry)  
XX DE Enterococcus faecalis DNA for cellular proliferation protein #619.  
XX KW Antisense; ds; prokaryotic cellular proliferation gene;  
XX KW antibiotic; antibacterial; drug design.  
XX OS Enterococcus faecalis.  
XX PN WO200170955-A2.  
XX PD 27-SEP-2001.  
XX PR 21-MAR-2001; 2001WO-US09180.

```
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX (ELITR-) ELITRA PHARM INC.
XX
XX Haasebeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
PI Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
XX P-PSDB; AAU35332.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Claim 27; Seq ID No 6828; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 867 BP; 266 A; 145 C; 209 G; 247 T; 0 other;
SQ
XX
XX Query Match 8.2%; Score 121; DB 23; Length 867;
XX Best Local Similarity 53.0%; Pred. No. 4.7e-24;
XX Matches 259; Conservative 0; Mismatches 230; Indels 0; Gaps 0;
```

```
DB 688 ATCGTTTGTGCGCCCGCGGTATGAAACAACGATTCGTCAAGAGTCCGAGATGAT 747
OY 598 GTGACAGACGGCGAGAAATTTGGTCAAACTGCTGTATGTAATTTGTCGCCACTC 657
DB 748 TTTCAAAAGCCGAGTTTCTTTAGAACATGTTTGTAGATCAGATTGGCCAAAGAAAT 807
OY 658 CAATTGGCT 666
DB 808 CTTTGGCT 816
XX
XX RESULT 12
XX AAU13395
XX ID AAU13395 standard; DNA; 6021 BP.
XX
XX AAU13395;
XX
XX 19-MAR-1999 (first entry)
XX
XX Enterococcus faecalis genome contig SEQ ID NO:458.
XX
XX Enterococcus faecalis; contig; detection; Enterococcal infection;
XX vaccine; attenuation; computer readable medium; ds.
XX
XX Enterococcus faecalis.
XX
XX MO9850555-A2.
XX
XX 12-NOV-1998.
XX
XX 04-MAY-1998; 98WO-US08985.
XX
XX 14-NOV-1997; 97US-0066009.
XX 06-MAY-1997; 97US-0044031.
XX 16-MAY-1997; 97US-0046655.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Dillon PJ, Kunsch CA;
XX
XX WPI; 1999-045171/04.
XX
XX New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
XX Claim 1; Page 1691-1694; 2084pp; English.
XX
XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAU13398 to AAU13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
XX Sequence 6021 BP; 1920 A; 1007 C; 1415 G; 1673 T; 6 other;
SQ
XX
XX Query Match 8.2%; Score 121; DB 20; Length 6021;
XX Best Local Similarity 53.0%; Pred. No. 1.1e-23;
XX Matches 259; Conservative 0; Mismatches 230; Indels 0; Gaps 0;
```

Db 5321 GATGAAGCTGTCTTAAACGGAGAACAAATTAAGAGCAACAGCTGTGCAATTTGGAAATC 5380  
 QY 238 TCCGATTTTCTTCTCCGCGGTTCTTTGGACAGCGTGGCGGATCATGAAAG 297  
 Db 5381 ATGGAAGCTTAATTTATCATAGGCAATATGGAAACATTTTGTGTAATAAATCACAGC 5440  
 QY 298 GCGATTACCGCGCCACAGAGCTGAAATCTCCCATGCTGTCTCCCTGCTTCGGTGTG 357  
 Db 5441 TTGTTAGAGGGGCGACAGAAAAGCATTTACAGTAGTAGTTTCACTGATCTGGTGTG 5500  
 QY 358 GCGCGATGAGAAACATCAAGCTTTGTATGATGTGTCAATACCGCGCTGTG 417  
 Db 5501 GCCGATATCAAGAAAGAAATTTTTCATTTGATCAATATGGCGAAAATTTGCGCGCTTG 5560  
 QY 418 CAGCGTACCGAGGCGCATTTGCGGTTCTGTGATTTTGGCAATCCAGATGGT 477  
 Db 5561 CAAAGGCAATACAAAGCAGGCTGTGATCTTAACGATATGCTGATCCAAAGACTGTG 5620  
 QY 478 GCGCGCATGCGCTGTGCGGTTCTATCTGAGCATCTCATTTTGGCAACCGCGCGCAG 537  
 Db 5621 GGTGTACCGCAAGTTTGGATGATGGAGATATTTTGGCAAGCCTCAGATTTA 5680  
 QY 538 ATAGGTTTCTGTGGTCTTGGCGGTGTGAGTTAACACTGGCGATGCGCTTCAGACGGT 597  
 Db 5681 ATCGGTTTGTGCTGGCGCGCGGTGAATGAACAAACGATTCGTCAAGATTGCCAGATGAT 5740  
 QY 598 GTGACAGCAGCGAGAAATTTGTGAAACCTGGTGTGATTTGATGAAATTTGTGTGCCATCTC 657  
 Db 5741 TTTCAAAAGCGCAGATTTCTTTAGAACATGTTTGTGATCAGATTGTGCAAGAAAT 5800  
 QY 658 CAATTGCGT 666  
 Db 5801 CTTTTCGCT 5809  
 RESULT 13  
 ABS99190  
 ID ABS99190 standard; DNA; 6021 BP.  
 AC ABS99190;  
 DT 18-DEC-2002 (first entry)  
 DE Enterococcus faecalis contig sequence #458.  
 XX  
 XX Computer readable medium; Enterococcus faecalis; microbe; growth;  
 KM pathogenicity; vaccine; resistance; Enterococcal infection; commercial;  
 KM therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;  
 KM biotech technology; antibacterial; modulator of nucleic acid expression;  
 KM contig; ds.  
 OS Enterococcus faecalis.  
 XX  
 XX US2002120116-A1.  
 PD 29-AUG-2002.  
 XX  
 XX 04-MAY-1998; 98US-0070927.  
 PF 04-MAY-1998; 98US-0070927.  
 XX  
 XX 04-MAY-1998; 98US-0070927.  
 ER  
 XX (KUNS/) KUNSCH C A.  
 PA (DILL/) DILLON P J.  
 PA (BARA/) BARASH S.  
 XX  
 XX Kunsch CA, Dillon PJ, Barash S;  
 FI  
 DR WPI; 2002-750065/81.  
 XX  
 XX Computer readable medium having recorded on it a Enterococcus faecalis  
 PT nucleotide sequence useful for detecting diseases related to  
 KM Enterococcus infections in animals  
 XX

PS Claim 1; Page -; 119pp; English.  
 CC The present invention relates to a new computer readable medium with an  
 CC Enterococcus faecalis nucleotide sequence. The invention is useful to  
 CC diagnose the presence of E. faecalis in a sample or determining the  
 CC presence of a specific microbe in a sample. The invention is also useful  
 CC for modulating the growth or pathogenicity of E. faecalis, in a vaccine  
 CC to confer resistance to Enterococcal infection, for commercial,  
 CC therapeutic and industrial purposes, and for fermenting a particular  
 CC sugar source or to produce a particular metabolite. The invention is  
 CC useful for detecting diseases related to Enterococcus infections in  
 CC animals, and for detecting E. faecalis using biotech technology. The  
 CC present nucleic acid sequence represents an Enterococcus faecalis contig  
 CC DNA sequence of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification but was obtained in electronic format directly  
 CC from USPTO at <http://seqdata.uspto.gov>.  
 XX  
 SQ Sequence 6021 BP; 1920 A; 1007 C; 1415 G; 1673 T; 6 other;  
 Query Match 8.2%; Score 121; DB 24; Length 6021;  
 Best Local Similarity 53.0%; Pred. No. 1.1e-23;  
 Matches 259; Conservative 0; Mismatches 230; Indels 0; Gaps 0;  
 QY 178 GATGAAGCTGTCTTAAACGGAGAACCAACCGTGGAGGCGGATTCGGTACCTTATTTG 237  
 Db 5321 GATGAAGCTGTCTTAAACGGAGAACCAATTAAGAACAGCTGTGCAATTTGCAATC 5380  
 QY 238 TCCGATTTTCTTCTCCGCGGTTCTTTGGACAGCTGCGCTGCGCATGAAAG 297  
 Db 5381 ATGGAAGCTTAATTTATCATAGGCAATATGGAAAGATTTGTGTAATAAATCACAGC 5440  
 QY 298 GCGATTACCGCGCCACAGAGCTGAAACTCCACCTGCTGTCTCCCTGCTTCGGTGTG 357  
 Db 5441 TTGTTAGAGGGGCGACAGAAAAGCATTTACAGTAGTAGTTTTCATCTGATCTGTGCT 5500  
 QY 358 GCGCGATGAGAAACATGAGCTTTTGTGATGATGTGTCAATACCGCGCGCTGTG 417  
 Db 5501 GCCGATATGCAAGAAAGAAATTTTTCATTTGATGCAAAATGGCAAAATTTGCGCCTTTG 5560  
 QY 418 CAGCGTACCGAGGCGCATTTGCGGTTCTGTGATTTTGGCAATCCAGATGGT 477  
 Db 5561 CAAAGGCAATACAAAGCAGGCTGTGATGATGCAATGATGCAATGCAAGACTG66 5620  
 QY 478 GCGCGCATGCGCTGTGCGGTTCTATCTGAGCATCTCATTTTGGCAACCGCGCGCAG 537  
 Db 5621 GGTGTACCGCAAGTTTGGATGATGGCAATATTTTGGCAGAGCCTCAGAGTTTA 5680  
 QY 538 ATAGGTTTCTGTGGTCTTGGCGGTGTGAGTTAACACTGGCGATGCGCTTCAGACGGT 597  
 Db 5681 ATCGGTTTGTGCTGGCGCGCGGTGAATGAACAAAGATTCGTCAAGATTGCCAGATGAT 5740  
 QY 598 GTGACAGCAGCGAGAAATTTGTGAAACCTGGTGTGATTTGATGAAATTTGTGTGCCATCTC 657  
 Db 5741 TTTCAAAAGCGCAGATTTCTTTAGAACATGTTTGTGATCAGATTGTGCAAGAAAT 5800  
 QY 658 CAATTGCGT 666  
 Db 5801 CTTTTCGCT 5809  
 RESULT 14  
 ABX65750  
 ID ABX65750 standard; DNA; 676 BP.  
 AC ABX65750;  
 DT 07-MAY-2003 (first entry)  
 DE Helicobacter pylori selected interacting domain (SID) DNA #349.  
 KM Protein-protein interaction; ulcer; selected interacting domain;  
 KM SID; gene; ds.  
 XX

```

XX Helicobacter pylori.
OS
XX WO200266501-A2.
XX
XX 29-AUG-2002.
XX
XX 28-DEC-2001; 2001WO-EP15428.
XX
XX 02-JAN-2001; 2001US-259302P.
XX
XX (HYBR-) HYBRIGENICS.
XX (INST) INST PASTEUR.
XX
XX Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
XX WPI; 2002-674910/72.
XX DR P-PSDB; ABU51006.
XX
XX New complexes of protein-protein interactions in Helicobacter pylori,
XX useful for identifying modulating compounds for treating or preventing
XX ulcers in mammals -
XX
XX Claim 7; Page 172; 642bp; English.
XX
XX The invention describes a complex of protein-protein interactions in
XX Helicobacter pylori selected from 421 complexes given in the
XX specification. The complex of protein-protein interactions are useful
XX for screening for agents which modulate the interaction of proteins.
XX Modulating compounds which binds to a targeted bacterial protein may be
XX used for treating or preventing ulcers in a human or animal. This
XX sequence encodes a selected interacting domain (SID), identified via
XX protein-protein interactions.
XX
XX Sequence 676 BP; 183 A; 119 C; 190 G; 184 T; 0 other;
XX
XX Query Match      8.1%; Score 119.2; DB 24; Length 676;
XX Best Local Similarity 50.9%; Pred. No. 1.4e-23;
XX Matches 283; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
XX
OY 93 TCTTGGATGAAATCCCAATATGACAACTCAATCAAGGCTATGAGACCTTGA 152
DB 73 TTTAGGGCTTAATGATCTTTAAATTTGGTGATTAAGAGAGCTATTAACAACGATTA 132
OY 153 GGGGGCTGGAAGCAAGGCCAATGCGATGATTCGGTAATTAATGAGAAAGCAACGTGA 212
DB 133 AAAATAGCAAAAAGAGACTTAACCGCCAAAGCTCAGTATGACGGGTGAGGCTAAATCA 192
OY 213 GGGGATTCGGGTAGCCGTTATTTGTCGATTTTTCCTTCTCGGCGGTTCTTTGGGAC 272
DB 193 CCGCATGCTTTGACAGATGCTGTGTTGATTTTACCTTTATGGGGGAGATTAGGCTC 252
OY 273 GGTGCGGTGCGTGCATGATGAAAGCGATTCACCGCGCACAGAGCTGAAATCCCACT 332
DB 253 TGTGAGAGGCGAAAAGATGTAAGAGCAATCATGCGCGGCTGCTAAAGAGAACGCTT 312
OY 333 GGTGCTTCCTCTGCTTCCGCTGCTGCGGCGATGACAGAAAGCAATCCAGCTTTGTAT 392
DB 313 ATTGATTTGTTTACGAGAGTGAGGGGGGCTAGATGAGAAAGATCACTTTATTCCTATGA 372
OY 393 GATGATGTCATTAACCGGGCTGTGAGCGTCACGCGAGGCGCAATTTGCCGTTCTGTGT 452
DB 373 AATGCTTAAACGAGAGCGGGCTTTGACCGATTTGAGTAGGGCAAACTCCCTTTATTTTC 432
OY 453 GTATTTGGCAATCCCAAGATGAGTGCGCAATGCGCTTGCGGGGTTCAATCGGCGATCT 512
DB 433 GCTTTAAGCGATCCCACTTATGGGGGCGTTAGCGGATCTTTTGCTTTTATGGGAGATCT 492
OY 513 CACTTTTGGGGAACCGCGCGCGCAATAGTTTCTCGGCTCTCGCGTGTGTGAGATTAC 572
DB 493 CATTATCGAGAGCCAGAGGGCGATGATGCTTTTGGGGGCGCTTAGGCTATTAACCAAC 552
OY 573 CACTGGGATGCGCTTCCAGACGAGTGTCAGAGGCGGAGAAATTTGGTAAATCTGTGT 632

```

```

DB 553 TATAGGGCGGATTTGCTGAGGGCTTCAACAGCGGAATTTTATTAAGCATGCTT 612
OY 633 GATTGATGAATTGCTG 648
DB 613 GATTGATGATGATGCTG 628

RESULT 15
AAS53651
ID AAS53651 standard; DNA; 870 BP.
AC AAS53651;
XX
XX 13-FEB-2002 (first entry)
XX
XX Helicobacter pylori DNA for cellular proliferation protein #105.
XX
XX Antisense; ds; prokaryotic cellular proliferation gene;
XX antibiotic; antibacterial; drug design.
XX
XX Helicobacter pylori.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELITRA) ELITRA PHARM INC.
XX
XX Haeelbeck R, Ohlsen KU, Zykind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
XX DR P-PSDB; AAU35792.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Claim 27; Seq ID No 7288; 511bp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence encodes an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 870 BP; 253 A; 147 C; 229 G; 241 T; 0 other;
XX
XX Query Match      8.1%; Score 119.2; DB 23; Length 870;

```

Best Local Similarity 50.9%; Pred. No. 1.5e-23;  
Matches 283; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

```
QY 93 TTCTTGGAAATGAAATCCCAATATATGCAACCTCAATCAAGCTATGCAAGACCTTGGG 152
   |||||
Db 240 TTACGACCTTAATGATCTTTAAATTTCTGTGATTAAGAGACTATTAACAACGCAATTA 299
   |||||
QY 153 GCGGGCTCGAAGCAAGGCCAATATCGATGATCGTAAATTAATGAGAAAGCACCGTGG 212
   |||||
Db 300 AAAATAGAAAAAGAAAGCTAAACGCCCAACTAGTATAGCGGTGAGGCTAAATCA 359
   |||||
QY 213 GGGCAATCCGGTACCGTTATTTGTCCGATTTTTCCTTCCTCGCGGTTCTTTGGGAC 272
   |||||
Db 360 CCGCATGCTTTGCAATCGTGTGTTGATTTTATGCTTATGAGGGGAGTTTAAAGCTC 419
   |||||
QY 273 GGTGCGGTGGGTGCGGATCATGTAAGGCGATTACCGGCCACAGAGCTGAACTCCCACT 332
   |||||
Db 420 TGTGGAGGGCGAAAAAGATGTAAGACCAATCAATCGCGGTGCTAAAGAGAACGTT 479
   |||||
QY 333 GCTGTCTCCCTCTTCCGCTGTGCGGATGCGGATGCAAGAAACATCGAGCTTTGTCA 392
   |||||
Db 480 ATTGATTGTTTGAAGAGAGTGGGGGCTAGGATGCAAGAAATCACTTATTCGCTCATGCA 539
   |||||
QY 393 GATGTGTCCATTAACCGCGGCTGTGACGCTACCGGAGGCGCATTTGCCGTTCTTGGT 452
   |||||
Db 540 AATGGCTAAACGAGGGCGGCTTTGAACGATGAGTGAAGGCCAACTCCCTTTCATTTTC 599
   |||||
QY 453 GTATTTCGCAATCCCAAGCATGAGTGGCGCATGCGCTCGTGGGTTCACTGGGCACTT 512
   |||||
Db 600 GCTCTTAAGCGATCCCACTTAATGGGGCGTTAGCGCATCTTTTGTCTTTTAAAGGGATCT 659
   |||||
QY 513 CACTTTTGGCGAACCAGCGCGCAGATAGTTTCTGTGGTCTCGCGTGTGAGTTTAA 572
   |||||
Db 660 CATTAATCGAAGCCAGGGGCGATGATAGGCTTTGGGGGCTTAGGGTATTAAGCAAC 719
   |||||
QY 573 CACTGGGCATGCGCTTCCAGACGCTGTGACAGCGCGAGAAATTTGGTGAACCTGTGT 632
   |||||
Db 720 TATAGGGGCGGATTTGCTGAGGGCTTTCAACAGCGGAAATTTTATAGACATGCGCTT 779
   |||||
QY 633 GATTGATGGAATTTGTG 648
   |||||
Db 780 GATTGATGATTTGTG 795
   |||||
```

Search completed: November 12, 2003, 20:26:11  
Job time : 440.983 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 18:57:58 ; Search time 3242.57 Seconds  
(without alignments)  
11040.790 Million cell updates/sec

Title: US-10-024-370-2

Perfect score: 1473

Sequence: 1 gtgagaagcgtttccgac.....gatttcagcgtttcagcag 1473

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_ges\_hum:\*  
18: em\_ges\_inv:\*  
19: em\_ges\_pln:\*  
20: em\_ges\_vrt:\*  
21: em\_ges\_fun:\*  
22: em\_ges\_mam:\*  
23: em\_ges\_mus:\*  
24: em\_ges\_pro:\*  
25: em\_ges\_rtd:\*  
26: em\_ges\_phg:\*  
27: em\_ges\_vrl:\*  
28: gb\_ges1:\*  
29: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88	6.0	488	28	BH379574 AG-ND-133
2	86.2	5.9	707	13	BH379574 AG-ND-133
3	84.2	5.7	604	9	BH379574 AG-ND-133
4	82.6	5.6	773	12	BH379574 AG-ND-133

Result No.	Score	Query Match	Length	ID	Description
5	81.2	5.5	1063	29	BH379574 AG-ND-133
6	81	5.5	712	12	BH379574 AG-ND-133
7	79.8	5.4	580	14	BH379574 AG-ND-133
8	79.8	5.4	792	10	BH379574 AG-ND-133
9	78.8	5.3	1042	29	BH379574 AG-ND-133
10	78.4	5.3	539	12	BH379574 AG-ND-133
11	77.6	5.1	652	29	BH379574 AG-ND-133
12	75	5.1	617	14	BH379574 AG-ND-133
13	74	5.0	668	9	BH379574 AG-ND-133
14	73.8	5.0	677	28	BH379574 AG-ND-133
15	73.6	5.0	606	14	BH379574 AG-ND-133
16	73.4	5.0	453	14	BH379574 AG-ND-133
17	73.4	5.0	708	28	BH379574 AG-ND-133
18	72.2	4.9	675	28	BH379574 AG-ND-133
19	72.2	4.9	696	28	BH379574 AG-ND-133
20	72.2	4.9	700	28	BH379574 AG-ND-133
21	72.2	4.9	703	28	BH379574 AG-ND-133
22	72.2	4.9	704	28	BH379574 AG-ND-133
23	72.2	4.9	709	28	BH379574 AG-ND-133
24	72.2	4.9	711	28	BH379574 AG-ND-133
25	72.2	4.9	752	28	BH379574 AG-ND-133
26	72.2	4.9	781	28	BH379574 AG-ND-133
27	72.2	4.9	807	28	BH379574 AG-ND-133
28	72.2	4.9	844	28	BH379574 AG-ND-133
29	71.8	4.9	1872	28	BH379574 AG-ND-133
30	71.4	4.8	607	14	BH379574 AG-ND-133
31	71.4	4.8	637	28	BH379574 AG-ND-133
32	70.4	4.8	843	28	BH379574 AG-ND-133
33	70	4.8	719	28	BH379574 AG-ND-133
34	69.4	4.7	697	28	BH379574 AG-ND-133
35	69	4.7	405	28	BH379574 AG-ND-133
36	68.8	4.7	627	9	BH379574 AG-ND-133
37	68.4	4.6	748	28	BH379574 AG-ND-133
38	68.4	4.6	792	28	BH379574 AG-ND-133
39	68.2	4.6	523	28	BH379574 AG-ND-133
40	68	4.6	698	28	BH379574 AG-ND-133
41	67.8	4.6	452	9	BH379574 AG-ND-133
42	67.4	4.6	722	28	BH379574 AG-ND-133
43	67	4.5	886	29	BH379574 AG-ND-133
44	66	4.5	1333	29	BH379574 AG-ND-133
45	65	4.4	558	12	BH379574 AG-ND-133

## ALIGNMENTS

RESULT 1  
LOCUS BH379574  
DEFINITION AG-ND-133C23, TF ND-TAM Anopheles gambiae genomic clone AG-ND-133C23  
ACCESSION BH379574  
VERSION BH379574.1 GI:17325716  
KEYWORDS GSS.  
SOURCE Anopheles gambiae  
ORGANISM Anopheles gambiae (African malaria mosquito)  
REFERENCE 1 (bases 1 to 488)  
AUTHORS Hong, Y.S., Hogan, J.R., Wang, X., Sarkar, A., Sim, C., Loftus, B.J., Ren, C., Huff, E.R., Carille, J.L., Black, K., Zhang, H.-B., Gardner, M.J.  
TITLE Construction of a BAC library and generation of BAC end sequence-tagged connectors for genome sequencing of the African malaria mosquito Anopheles gambiae  
JOURNAL Mol. Genet. Genomics 268 (6), 720-728 (2003)  
MEDLINE 22542063  
PUBMED 12655398  
COMMENT Other\_GSSs: AG-ND-133C23.TR  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjoftus@igr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 For

# FEATURES

## source

1. .488  
/organism="Anopheles gambiae"  
/mol\_type="genomic DNA"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-13C23"  
/clone\_1ib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 137 a 86 c 109 g 156 t

## ORIGIN

Query Match 6.0%; Score 88; DB 28; Length 488;

Best Local Similarity 49.0%; Pred. No. 7.2e-13;

Matches 235; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

132 AGGCTATCAGAGACCTTGGAGCGGCTCGAAGGCAAGGCAATGATGATCGGTAT 191  
1 AGCTTATCTGACCGATTAAAGAGGTAAAGGCAAGGCAATGATGATGATCGGTAT 60  
192 TACTGAGAAGGCAAGGCAAGGCAATGATGATGATGATGATGATGATGATGAT 251  
61 TATGCTGTGTGTAAAGTAAAGGCAAGGCAATGATGATGATGATGATGATGAT 120  
252 CCTGCGCGGCTTGTGGGACCGTGGCGGTGGTGGCATGAGAGGCAATGATGAT 311  
121 TATCGAGAGGCTATGATGCTGTGATGAGGAGGAGGAGGAGGAGGAGGAGGAG 180  
312 CACAGAGTGAAGTCCGACGCTGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 371  
181 AATGAGATGAATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 240  
372 AGACATGAGAGCTTTGTCATGATGATGATGATGATGATGATGATGATGATGAT 431  
241 AGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
432 GCGGCAATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 491  
301 CAAGGCTTACCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 360  
492 GTGGGCTTATGAGGCAATGATGATGATGATGATGATGATGATGATGATGATGAT 551  
361 TTTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
552 TCTTCGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 611  
421 TCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480

## RESULT 2

BUE54446

LOCUS BUE54446 707 bp mRNA linear EST 30-SEP-2002

DEFINITION 1112131P06.y1 C. reinhardtii CC-1690 (mt+), CC-1691 (mt-), Gamete

(normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA

SEQUENCE

BUE54446

KEYWORDS

EST. BUE54446.1 GI:23366627

SOURCE Chlamydomonas reinhardtii

ORGANISM Chlamydomonas reinhardtii

REFERENCE Eukaryote: Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

AUTHORS Chlamydomonadaceae; Chlamydomonas.

1 (bases 1 to 707)

TITLE Grosman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre

, P., McDermott, J.P., Shlager, J., Sifflow, C. and Stern, D.

Analyses of the Chlamydomonas reinhardtii Genome: A Model,

Unicellular System for Analyzing Gene Function and Regulation in

Vascular Plants. Project: 1112

Unpublished

JOURNAL Contact: Charles Hauser

COMMENT DCMB Box 91000

Duke University

Durham, NC 27708-1000

Tel: 919 613 8159

Fax: 919 613 8177

Email: chauser@duke.edu.

# FEATURES

## source

1. .707  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="21gr (CC-1690 wild type mt+) & 6145C (CC-1691  
wild type mt-)"  
/db\_xref="taxon:3055"  
/clone\_1ib="C. reinhardtii CC-1690 (mt+), CC-1691 (mt-),  
Gamete (normalized), Lambda Zap II"  
/note="Vector: pLuscript II SK-; Site\_1: EcoRI; Site\_2:  
XhoI; Gamete library was constructed by Hui Zhao, Min Lu,  
Jeffrey McDermott, William J. Snell and John Davies.  
Strain 21gr cells (CC-1690; mating type plus) and strain  
6145C cells (CC-1691; mating type minus) that had been  
growing on a light-dark cycle (13:11 L/D) in R-medium  
(Sager and Granick) were separately transferred into  
nitrogen-free medium at 8 hours into the light period.  
PolyA mRNA was purified from each sample every 2 hours for  
the next 18 hours. The mRNA was pooled and used for cDNA  
synthesis. The cDNA was directionally cloned into lambda  
Zap II (Stratagene) in the EcoRI (5') and XhoI (3')  
sites. pLuscript II SK- plasmids were excised from the  
lambda Zap clones by superinfection with Exsist  
(Stratagene) phage. The library was normalized using  
method 4 described in Bonaldo et al., (1996) Genome  
Research 6: 791-806."

BASE COUNT 127 a 241 c 227 g 112 t

## ORIGIN

Query Match 5.9%; Score 86.2; DB 13; Length 707;

Best Local Similarity 50.6%; Pred. No. 2.5e-12;

Matches 208; Conservative 0; Mismatches 203; Indels 0; Gaps 0;

942 CAGCGTGGCGCGCAGAGAGCTGCTTTGGCGCTGCGGCAATTCGCTGGCGCGAGCT 1001  
286 CATGCCCGCAGCCCAAGCGCTACCCCAAGGCGCTGCGCTTACATGCGCCAGCAAGTT 345  
1002 AAACCTGCGCATCGTGCATCATGACACCTCCGCGCGGCAATTGTCGAGCGCGCTGA 1061  
346 TGTGTGCGCATCATCATCTTGTGAGACGCCCGGAGCCTATGCCGCAAGACCGGGA 405  
1062 GAGAGCTGGCATCCGCAAGCTGATGCGCGACCTTGTCCAGATTATGACGCTCCCT 1121  
406 GAGAGCTGGCGCAGGCGCAGGCGCATTCGCGGAACTCGGAGATGTTCCGCTCGCGCT 465  
1122 CCCACGCTTTCGCTATTTATGTCAGGCGCTGGCGGCGCGCTGCGCATGCTGCC 1181  
466 GCCCATCATCTCGGTGTGTCATTTGCGAGGCGCGCTCGCGCGCGCTGCGCATTTGCTG 525  
1182 CGCCGATCTGTCTACCGCGCGGCAAGCGCTGCTTCGCGCATTCGACAGGCGGCGC 1241  
526 CGCCACCGCAACTGTATCATGAGGAGACGCGTCTACTAGTGGCTCGCGCGGCGCTG 585  
1242 CTGCGCATCTCTTTCGCGCAGACCAACGCGCGGCAATCATAGAGCGACAGGCGT 1301

Db 586 CGCCGCATCTGTGGAAGACGCTCTGCGCCGCGGAGGCCACTGAGCCCTCGCAT 645  
 QY 1302 GCAGGCGACGACCTTTTAAGCCAGGGCTTATCGACGGATCGTCCGCA 1352  
 Db 646 CACCTCGGCGGAGCTGTGTAAGTTGCGCTCATGACACATCGTCCGCA 696

RESULT 3  
 AM584783  
 LOCUS N2109176 MHAM Medicago truncatula/Glomus versiforme mixed EST  
 DEFINITION  
 ACCESSION AM584783  
 VERSION AM584783.1 GI:7261837  
 KEYWORDS  
 SOURCE Medicago truncatula/Glomus versiforme mixed EST library  
 ORGANISM Medicago truncatula/Glomus versiforme mixed EST library  
 Eukaryota; mixed EST libraries.  
 1 (bases 1 to 604)  
 Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D.,  
 Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.  
 ESTs from roots of Medicago truncatula after colonization with  
 Glomus versiforme  
 JOURNAL Unpublished  
 COMMENT Contact: Harrison M.J.  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73401  
 Tel: 580-223-5810  
 Fax: 580-223-7380  
 Email: mjharrison@noble.org  
 Other name: MHAM-7b-E05; Date: 3/14/00; Updated to the database of  
 Expressed Sequence Tags (dbEST) on 04/27/00; More information is  
 available at <http://chryslie.tamu.edu/medicago>.  
 Seq primer: 73.

FEATURES  
 source  
 1..604  
 Location/Qualifiers  
 /organism="Medicago truncatula/Glomus versiforme mixed EST  
 library"  
 /mol\_type="mRNA"  
 /cultivar="Medicago truncatula genotype A17"  
 /db\_xref="taxon:119092"  
 /clone="MHAM-7110"  
 /tissue\_type="roots colonized with Glomus versiforme"  
 /dev\_stage="Roots harvested at 10, 17, 22, 31 and 38 days  
 post-inoculation with Glomus versiforme. The library was  
 made from a mixture of RNA from each of these stages."  
 /lab\_host="E. coli strain XL0LR"  
 /clone\_lib="MHAM"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; cDNA was prepared from polyA+ enriched RNA from  
 roots harvested at 10, 17, 22, 31 and 38 days  
 post-inoculation with Glomus versiforme. The cDNA was  
 directionally ligated into the UniZap XR vector from  
 StrataGene and packaged using Gigapack III Gold packaging  
 extracts. Plasmids containing cDNA inserts were excised  
 from the recombinant lambda-Zap phage using Ex-aseIst  
 helper phage and propagated in XL0LR cells."

BASE COUNT 179 a 95 c 136 g 194 t  
 ORIGIN

Query Match 5.7%; Score 84.2; DB 9; Length 604;  
 Best Local Similarity 48.8%; Pred.No. 8.3e-12;  
 Matches 258; Conservative 0; Mismatches 268; Indels 3; Gaps 1;

QY 127 AATCAAGCTATGACAGACCTTGAAGCGGCTCGAAGCAAGCCAAATGCATGATCG 186  
 Db 13 AATGATGATTAACAAGATGCTTGAATTTCTTCAAGACAGAACCGATTAAGATCG 72  
 QY 187 GTAATTACTGGAAGACCGCTGAGGGGCAATTCGGGTAGCCGTATTTTGTCCATTTT 246  
 Db 73 GTTCAACAGGACGACGCTCAAGTAATGATTAATCTGTAGCAATAGTATTAATGATTTT 132

QY 247 TCCTTCCTCGGCGGTTCTTTGGGACAGCTCGCTGCTGTCGCATCATGAAAGCGATTGAC 306  
 Db 133 GAGTTTATGGAGAGTATGAGATCCGATAGGGGTGAGAAATACATCTCGTTGATTGA 192  
 QY 307 CGCGCACAGAGCTGAAACTCCACTGCTGTCTCCCTGTCGGTGGTGGCGCATG 366  
 Db 193 TATGCTACCAATCAACGTTTACTCTTATTAATGATGGCTGAGGAGCGCGATTG 252  
 QY 367 CAGAGACCAATTCAGCTTTTGTATGATGATGTCATTAACCGGCGCTGACGCTGAC 426  
 Db 253 CAGAGAGAGAGTTGAGCTTATGCAAAATGCTAAATTTCTCTTTATATATATAT 312  
 QY 427 CGCGAGCGCATTTGCCCTTCCTGCTGATTTGCGCA--ATCCACAGATGGTGGCCG 483  
 Db 313 CAATCATCAAAAGTATTTCTATGTACCGATCTTACATCTCCATCACTAGTGGGTA 372  
 QY 484 ATGGCTTGCGGGGTTCAATCGGCAATCTATTTGGGAAACCGGCGCGCATAGGT 543  
 Db 373 ACAGCTAGTTTGGATGTGGGGGATATCAATTCCTGCGGAACCGATCTTACATGCA 432  
 QY 544 TTCTCGGCTCTCGCGTGTGAGTATCACTGCGCATGGCGCTTCAGACGCTGTGAG 603  
 Db 433 TTGCAAGTAAAGATGATTAAGAAAGCTTGAAGATTGAAGTCCGAAAGTATACAA 492  
 QY 604 CAGCGGAGAAATTTGTTGAATCTGTGATGATGATGATGATTTGTCGC 652  
 Db 493 TCGGCTGAATTTTATTCGAAAGAGTGATTTGATTCATCTGATCCAC 541

RESULT 4  
 B1311746  
 LOCUS EST5313496 GUSD Medicago truncatula cDNA clone pGSD15E20 5' end,  
 DEFINITION  
 ACCESSION B1311746  
 VERSION B1311746.1 GI:14986073  
 KEYWORDS  
 ORGANISM Medicago truncatula (barrel medic)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
 Medicago.  
 1 (bases 1 to 773)  
 Grusak,M.A., Sanac,D.A., Town,C.D., Van Aken,S., Utterback,T., Cho  
 ,J. and Fraser,C.M.  
 ESTs from developing reproductive tissues of Medicago truncatula  
 JOURNAL Unpublished  
 COMMENT Contact: Michael A. Grusak  
 USDA/ARS Children's Nutrition Research Center  
 Baylor College of Medicine  
 1100 Bates Street, Houston, TX 77030-2600, USA  
 Tel: 713-798-7044  
 Fax: 713-798-7078  
 Email: mgrusak@bcm.tmc.edu  
 B1398601e  
 TIGR sequence name: MTPA234TK  
 More information is available at: [www.medicago.org](http://www.medicago.org)  
 Seq primer: SKmod (CTA GAA CTA GAT CC).  
 Location/Qualifiers  
 1..773  
 /organism="Medicago truncatula"  
 /mol\_type="mRNA"  
 /cultivar="A17"  
 /db\_xref="taxon:3880"  
 /clone="pGSD15E20"  
 /tissue\_type="immature seeds"  
 /dev\_stage="Immature seeds, 11 to 19 days after  
 pollination"  
 /clone\_lib="GUSD"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; Immature seeds, collected from pods ranging in age  
 from 11 to 19 days after pollination, were harvested from

greenhouse-grown plants. Seeds were removed and separated from pod walls and were immediately frozen in liquid nitrogen. Seeds throughout the age range were pooled for mRNA extraction. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into the Uniap XR vector from Stratagene and packaged using GigaPack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLR cells."

BASE COUNT 261 a 112 c 154 g 246 t  
ORIGIN

Query Match 5.6%; Score 82.6; DB 12; Length 773;  
Best Local Similarity 48.6%; Pred. No. 2.5e-11;  
Matches 257; Conservative 0; Mismatches 269; Indels 3; Gaps 1;

127 AATCAAGGCTATGACAGACCTTGGAGCGGCTCCAGCAAGGCCAATGCGATGATCG 186  
127 AATGATGATTCATCAAAATCGCTTATCTTATCAAGACAGACCGGATTAAGTGG 92  
187 GTAATTACTGGAAGACCGTGGAGGCGATTCGGTGAACCGTTATTTTCCGATTTT 246  
93 GTTCAACAGGCGACGTCACAGTAATGATGATTCCTGAGCAATGATGATGATTTT 152  
247 TCCCTCCGCGGCTTCTTTGGGCAAGGCGGCGGCGGCGATGATGAAGCGATTCAC 306  
153 GAGTTATGAGGAGTATGAGGATCCGATGAGGAGTGAAGAAATACCTCGTTGATGAA 212  
307 CGCGCCAGAGCTGAACTCCACCTGCTGATCTCCCTGCTTCCGCTGAGTGGCGCATG 366  
213 TATGCTACCAATCAACGTTTACCTTATTAATGATGCGCTCGAGAGGCGGATG 272  
367 CAGAGAACATCGAGCTTTGTCTATGATGTGTCATACCGCGGCTGTGAGCGTAC 426  
273 CAAGAGGAAGTTGAGCTTATGCAAAATGCTAAATTTCTGCTTTATATATATAT 332  
427 CGCGAGGCGCATTTCCGCTCGTGTGATTTGGCCAA--TCCACAGATGGTGGCGCC 483  
333 CAATCAATCAAAAGTTATCTATCTATACCGATATCTTACCTACCTAGTGGGATA 392  
484 ATGCGCTCGGAGGCTTCACTTGGGATCTCACTTTTGGGAAACCGCGCGGAGATAGT 543  
393 ACAGTATGTTTGGATGTTGGGAGATATCATCTTGGCGAACCGAGTCTTACATAGA 452  
544 TTCTCGGCTCTCGGCTGTGAGTTAACCACTGGGATGCGCTTCCAGAGCGTGTGAG 603  
453 TTTCAGAGTAAAGATGATGAAAGAACTTGAAGATCGAAGTCCCGAAGGTATACA 512  
604 CAGGCGGAGATTTGGTGAAGAACTGTGATGATGATGATGATGATGATGATGATG 652  
513 TCGGCTGATTTTATTCGAAAGGATGATTTGATTCATCTGTACAC 561

RESULT 5  
B2548722 1063 bp DNA linear GSS 17-DEC-2002  
LOCUS B2548722  
DEFINITION pacel-60\_1377.s1 pacel-60 Pseudomonas aeruginosa genomic clone  
pacel-60\_1377, genomic survey sequence.  
ACCESSION B2548722  
VERSION B2548722.1 GI:27152303  
KEYWORDS GSS.  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.  
REFERENCE 1 (bases 1 to 1063)  
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.  
TITLE Whole-genome-sequence variation among multiple isolates of Pseudomonas aeruginosa library  
J. Bacteriol., (2002) In press  
COMMENT Contact: Chris K. Raymond

Genome Center  
University of Washington  
Box 352145, Seattle, WA 98105-2145, USA  
Tel: 2062216954  
Fax: 2066857244  
Email: craymond@u.washington.edu  
Class: shotgun.

FEATURES  
source

1..1063  
location/Qualifiers  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="1-60"  
/db\_xref="taxon:287"  
/clone="pacel-60\_1377"  
/clone\_1bp="pacel-60"  
/note="Clinical isolate 1-60 Whole genomic shotgun library."

BASE COUNT 202 a 316 c 335 g 209 t 1 others  
ORIGIN

Query Match 5.5%; Score 81.2; DB 29; Length 1063;  
Best Local Similarity 49.8%; Pred. No. 6.6e-11;  
Matches 206; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

942 CACGCTTGGCGCGAGAGCTGCGCTTTGGCGGCTGCGGATTTCCGTGGCGCGGAGCT 1001  
58 CATGCCCGCTCCGGAAGGCTATGCCAAGGCTGCTGCTGATGGAATGGCCGACGCTT 117  
1002 AAACCTGCCGATGCTGTCATCATGACACCTCCGCGCGGAATTTGCGAGCGGCTGA 1061  
118 CAAGATGCCATCTTCACTTATGACAGCGCGGCGCTTACCGGGAGATGATGCGGA 1177  
1062 GGAGCTGGGATGCGAAGCTGATTTGGCGGCGACCTTTCAGATTTACAGCTGCCCT 1121  
178 GGAACGGCGCGCAAGCGAGGCGATGCTGGAACCTGCGGATGCGCGGATGGAAGAC 237  
1122 CCCACGTTTCCGCTATTTATTTGATCAGGCGGTTGGGAGGCGCGCTGACATGCTGCC 1181  
238 GCCATATGCGCACCGTATGCGCGAGGCGGTTCCGCGCGCGCTGGCCATCGGTGT 297  
1182 CGCGATCTGTCTACCGCGCGGCAAAACGCGTGTCTGCGCATTTGCCACAGAGGCGC 1241  
298 CTGGACAAGTTGAACATGCTGCAATCTCACCTATTCGATGATCTGCGCGGAAGATG 357  
1242 CTGGCGCATCTCTTCCGCGCACACCAACCAACCGCGCGGAATCTATAGAGCAAGGCGT 1301  
358 CGCCTCATCTCTGGAAGACCGCGGAGAGGCGCGGAAAGCGCGGACGATGGCAT 417  
1302 GCAGCGCGACGACCTTTTAAAGCCAAAGGCTTATGAGGAGTGTGCGGATGCTGCCGAAAC 1355  
418 CACCGCGAGCGCTTGAAGGCTTGGGATGCTGCAAGGTATTCAGACGAC 471

RESULT 6  
B1310727 712 bp mRNA linear EST 20-JUN-2001  
LOCUS B1310727  
DEFINITION ESTB1312477 GSSD Medicago truncatula cDNA clone pGSSB074 5' end,  
mRNA sequence.  
ACCESSION B1310727  
VERSION B1310727.1 GI:14985054  
KEYWORDS EST.  
SOURCE Medicago truncatula (barrel medic)  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.  
REFERENCE 1 (bases 1 to 712)  
AUTHORS Grusak, M.A., Samac, D.A., Town, C.D., Van Aken, S., Uterback, T., Cho, J., and Fraser, C.M.  
TITLE ESTs from developing reproductive tissues of Medicago truncatula  
JOURNAL Unpublished  
COMMENT Contact: Michael A. Grusak

USDA/ARS Children's Nutrition Research Center  
Baylor College of Medicine  
1100 Bates Street, Houston, TX 77030-2600, USA  
Tel: 713-798-7044  
Fax: 713-798-7078  
Email: mgrusak@bcm.tmc.edu  
B397582e

TIGR sequence name: MTPAL60TK  
More information is available at: www.medicago.org  
Seq primer: SKmod (CTA GAA CTA gtc gat cc).

## FEATURES

source

1..712

/organism="Medicago truncatula"

/mol\_type="mRNA"

/culivar="A17"

/db\_xref="taxon:3880"

/clone="G8SD8J24"

/tissue\_type="immature seeds"

/dev\_stage="Immature seeds, 11 to 19 days after

pollination"

/clone\_lib="G8SD"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI; Immature seeds, collected from pods ranging in age

from 11 to 19 days after pollination, were harvested from

greenhouse-grown plants. Seeds were removed and

separated from pod walls and were immediately frozen in

liquid nitrogen. Seeds throughout the age range were

pooled for mRNA extraction. cDNA was prepared from polyA+

enriched RNA. The cDNA was directionally ligated into

the Uniap XR vector from Stratagene and packaged using

GigaPack III Gold packaging extracts. Plasmids containing

cDNA inserts were excised from the recombinant lambda-Zap

phage using Ex-assist helper phage and propagated in

XLOLR cells."

BASE COUNT 222 a 107 c 153 g 230 t

ORIGIN

Query Match 5.5%; Score 81; DB 12; Length 712;  
Best Local Similarity 48.4%; Pred. No. 6.5e-11;

Matches 256; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

127 AATCAAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGCCAAATGCGATGATCG 186  
128 AATGATGATATCAAAATGCTTATCTTATCAAGACAGAACTGGATTACTGATGG 122  
187 GTAATTACTGGAAGACCGCTGAGGCGATTCGGTACGCCGTTATTTGCCATTTT 246  
188 GTTCAAAACAGGACAGGTCACAGTAATGCTATCTGTAGCAATAGTATATGATTTT 182  
247 TCCTTCTCGGCGGTTCTTTGGGCAOGGTCGGTGGGCGCATGATGAAGCGCATTCAC 306  
183 GAGTTTATGGAGAGTATGGAGATCCGAGTGGGTGAAATAATCACTCGTGTATGAA 242  
307 CGCGCACAGAGCTGAACTCCCACTGCTGCTCCCTGCTTCGGTGGTGGCGCATG 366  
243 TATGTACCAATCAACGTTTACCTTATTAATAGATGCGCTCGAGAGACGGGTATG 302  
367 CAGGAAGCATGACGCTTTTGTCAATGATGTCATPACCGGCGCTGTGACGCGTCA 426  
303 CAGGAAGAGAGTTGAGCTTATGCAAAATGCTAAATTTGCTTTTATATATATAT 362  
427 CGCGAGGCGCATTTCCGCTCTGCTGATATTTGGGCA---TCCACAGATGGTGGCC 483  
363 CAATCAATCAAAATTTATTTATATACGATATCTTACCTCTCTACTACTGTTGGGTA 422  
484 ATGGCTCTGGGGTTCATCTGGGATCTCACTTTTGGGAAACCGGCGCGCAGATAGT 543  
423 ACAGTAGTTTGAATGTGGGGATATCATCTTTTGGGAAACCGGATGCTTACTATACA 482  
544 TTCTCGGCTCTCGCGTGTGAGTTAACACTGGGATGCGCTTCCAGACGCTGTGAG 603  
483 TTTCGAGGTAAAGAGTATATGAAAGAACTTGAAGATCGAAGTCCCGAAGGTATACA 542

OY 604 CAGCGGAGAAATTTGGTGAAGAACTGCTGTGATGATGAAATTTGCTCC 652  
Db 543 TCGGCTGAATTTTATTTGAAAAGGCTGATTTGATCTACTGTACAC 591

RESULT 7  
CA990607  
LOCUS  
DEFINITION  
ESR64415 G8SD Medicago truncatula cDNA clone G8SD-29C10, mRNA  
sequence.  
ACCESSION  
CA990607  
VERSION  
CA990607.1 GI:27523501  
KEYWORDS  
EST.  
SOURCE  
ORGANISM  
Medicago truncatula (barrel medic)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosid1; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
Medicago.  
1 (bases 1 to 580)

REFERENCE  
AUTHORS  
Grusak, M.A., Samac, D.A., Town, C.D., Van Aken, S., Utterback, T.,  
Cheung, F. and Fraser, C.M.  
TITLE  
More ESTs from developing reproductive tissues of Medicago  
truncatula  
JOURNAL  
Unpublished  
COMMENT  
Contact: Grusak, M.A.  
USDA/ARS Children's Nutrition Research Center  
Baylor College of Medicine  
1100 Bates Street, Houston, TX 77030-2600, USA  
Tel: 713 798 7044  
Fax: 713 798 7078  
Email: mgrusak@bcm.tmc.edu

TIGR sequence name: MTPBX17K  
More information is available at: www.medicago.org  
Seq primer: SKmod (CTA GAA CTA gtc gat cc).  
Location/Qualifiers  
1..580

## FEATURES

source

/organism="Medicago truncatula"

/mol\_type="mRNA"

/culivar="A17"

/db\_xref="taxon:3880"

/clone="G8SD-29C10"

/tissue\_type="immature seeds"

/dev\_stage="Immature seeds, 11 to 19 days after

pollination"

/clone\_lib="G8SD"

/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:

XhoI; Immature seeds, collected from pods ranging in age

from 11 to 19 days after pollination, were harvested from

greenhouse-grown plants. Seeds were removed and

separated from pod walls and were immediately frozen in

liquid nitrogen. Seeds throughout the age range were

pooled for mRNA extraction. cDNA was prepared from polyA+

enriched RNA. The cDNA was directionally ligated into

the Uniap XR vector from Stratagene and packaged using

GigaPack III Gold packaging extracts. Plasmids containing

cDNA inserts were excised from the recombinant lambda-Zap

phage using Ex-assist helper phage and propagated in

XLOLR cells."

BASE COUNT 171 a 88 c 129 g 192 t

ORIGIN

Query Match 5.4%; Score 79.8; DB 14; Length 580;  
Best Local Similarity 49.5%; Pred. No. 1.3e-10;

Matches 235; Conservative 0; Mismatches 237; Indels 3; Gaps 1;

OY 181 GAATCGGAATTTACTGGAAGAACCGTGGAGGCGATTCGGTAGCCGTTATTTGCC 240  
Db 2 GATGGGTTCAAGACAGGTCACAGTAATGATATCTTCTGTAACAATGATATATG 61  
OY 241 GATTTTCTCTCTCGGCGGTTCTTTGGGCAOGGTCGGTGGGCGATCATGAAGCG 300  
Db 62 GATTTTGAATTTATGGAGGTATGAGATCCGATGAGGTGAGAAATACCTCGGTTG 121

```

Oy 301 ATTACCCGCGCACAGAGCTGAACTCCACCTGCTGCTCCCTCTTCGGGTGTCG 360
Db 122 ATTGAATATGTCTACCAATCAACGTTTACCTTATTAATGATGATGCGCTGGAGAGCG 181
Oy 361 CGCATGACAGAAAGAACATCGAGCTTTTGTATGATGATGTCATTAACCGCGCTGTGAG 420
Db 182 CGTATGACAGAAAGAAAGTTTATGCTTAATGCAAAATGGCTAAATTTCTGCTTTTATAT 241
Oy 421 CGTACCCGCGAGCGCATTTGCGCTTCTGCTGATTTGCGCA--ATCCACGATGGGT 477
Db 242 AATTATCAATCAATCAAAAGTTATTTATGATGATCCGATATCACTCTTACTACTGCT 301
Oy 478 GCGCGCATGCGCTCGTGGGGTTCATCTGGGATCTCACTTTTGGGAAACCGCGCGGAG 537
Db 302 GGGGTAACAGCTAGTTTGGAAATGTTGGGGAATATCTTTCGGAACCGCATGCTTAC 361
Oy 538 ATAGGTTTCTTGCTGCTCGCTGCTGAGTTAACCACTGGGACATGCGCTTCACAGCGT 597
Db 362 ATAGCATTTGCGAGTAAAGAGTATTTGAGAGAAAGTTGAAGATGGAAGTCCGCAAGT 421
Oy 598 GTGACGACGCGGAGAAATTTGGTGAACCTGGTGTGATGATGAAATTTGTGCG 652
Db 422 ATACATCGGCTGAATTTTATTCGAAAAGGCGTGAATTCACCTGTAACAC 476

RESULT 8
BGS84921 792 bp mRNA linear EST 11-APR-2001
LOCUS EST84921 MHAM Medicago truncatula/Glomus versiforme mixed EST
DEFINITION library cDNA clone PMHAM-20E21 5' end, mRNA sequence.
ACCESSION BGS84921
VERSION BGS84921.1 GI:13599985
KEYWORDS
SOURCE Medicago truncatula/Glomus versiforme mixed EST library
ORGANISM Medicago truncatula/Glomus versiforme mixed EST library
REFERENCE 1 (bases 1 to 792)
AUTHORS Harrison,M.J., Liu,J., Town,C.D., Van Aken,S., Uterback,T., Cho,J.
and Fraser,C.M.
TITLE ESTs from roots of Medicago truncatula after colonization with
Glomus versiforme, 2001
JOURNAL Unpublished
COMMENT Contact: Harrison M.J.
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
Noble EST name: N380393e TIGR sequence name: MTDBW35TK More
information is available at: http://www.medicago.org
Seq primer: SKmod (CTA GAA CTA gtg gat CC).
FEATURES
SOURCE Location/Qualifiers
1..792
/organism="Medicago truncatula/Glomus versiforme mixed EST
library"
/mol_type="mRNA"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="PMHAM-20E21"
/lisue_type="roots colonized with Glomus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The library was
made from a mixture of RNA from each of these stages."
/lab_host="E. coli strain XLOLR"
/clone_1ib="MHAM"
/notes="Vector: pBluescript SK-, Site_1: EcoRI, Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA from
roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The cDNA was
directionally ligated into the Unizap XR vector from
Stratagene and packaged using Gigapack III Gold packaging

```

```

BASE COUNT 264 a 112 c 151 g 265 t
ORIGIN
Query Match 5.4%; Score 79.8; DB 10; Length 792;
Best Local Similarity 49.5%; Pred. No. 1,4e-10;
Matches 235; Conservative 0; Mismatches 237; Indels 3; Gaps 1;
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in XLOLR cells."

Oy 181 GAATCGGTAATTACTGGAAGAGCACCGTAGAGGCAATCCGATGCGTTATTTGTC 240
Db 33 GATCGGTTCAACAGGACAGGTCAGTCAATGATATTCCTGATGCAATAGATATATG 92
Oy 241 GATTTTCTTCTCGCGCGTTCCTTTGGGACAGGTGCGCTCGGTGCGCATCAAGAGCG 300
Db 93 GATTTTGAATTTTAAAGGAGGTAGATGGAATCCGTAATGGGTGAGAAATCACTCGTTG 152
Oy 301 ATTACCGCGCACAGAGCTGAAACTCCACTGCTGCTCCCTGCTTCGGGTGTCG 360
Db 153 ATTGAATATGTCTACCAATCAACGTTTACCTTATTTATGATGTCGCTGGAGAGCG 212
Oy 361 CGCATGACAGAAAGAACATCGAGCTTTTGTATGATGTCATTAACCGCGCTGTGAG 420
Db 213 CGTATGACAGAAAGAAAGTTTATGAGCTTATGCAAAATGGCTAAATTTCTGCTTTATAT 272
Oy 421 CGTACCCGCGAGCGCATTTGCGCTTCTGCTGATTTGCGCA--ATCCACGATGGGT 477
Db 273 AATTATCAATCAATCAAAAGTTATTTATGATGATCCGATCTTCACTCTACTACTGCT 332
Oy 478 GCGCGCATGCGCTCGTGGGGTTCATCTGGGATCTCACTTTTGGGAAACCGCGCGGAG 537
Db 333 GGGGTAACAGCTAATTTTGGAAATGTTGGAGATATCACTTTCGCAACCGATGCTTAC 392
Oy 538 ATAGGTTTCTTGCTGCTCGCTGCTGAGTTAACCACTGGGACATGCGCTTCACAGCGT 597
Db 393 ATACATTTGCGAGTAAAGAGTATTTGAGAGAAAGTTGAAGATGGAAGTCCGCAAGT 452
Oy 598 GTGACGACGCGGAGAAATTTGGTGAACCTGGTGTGATGATGAAATTTGTGCG 652
Db 453 ATACATCGGCTGAATTTTATTCGAAAAGGCGTGAATTCACCTGTAACAC 507

RESULT 9
BZ561247 1042 bp DNA linear GSS 17-DEC-2002
LOCUS pacs2-164_3167.y3 pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pacs2-164_3167, genomic survey sequence.
ACCESSION BZ561247
VERSION BZ561247.1 GI:27180878
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE 1 (bases 1 to 1042)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: ckraymond@u.washington.edu
Class: shotgun.
FEATURES
SOURCE Location/Qualifiers
1..1042
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"

```



/db xref="taxon:287"  
/clone lib="pac2-164.3167"  
/note="clinical isolate 2-164 whole genomic shotgun library."  
BASE COUNT 207 a 320 c 270 g 244 t 1 others  
ORIGIN

Query Match  
Best Local Similarity 50.3%; Pred. No. 2.9e-10; Length 1042;  
Matches 194; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 970 GCGGCTGCTGCGATTCGCTGCGCGCGAGCTAAACCTGCGCATCTGTCATCATCGAC 1029  
DB 110 GCGTGGCGGCTGATGTAATGCGCGAAGCGCTTCAAGATCCATCTTCACTTCAGC 169  
QY 1030 ACCCTCGGCGCGGATTCGCGAGCGCGCTGAGAGCTTGGCGATCGCAAGCTGATTCG 1089  
DB 170 AGCGCGCGCGCTACCGCGGCGCATTCGATGCCGAGAAAGCGGCGAGCGGATCGCC 229  
QY 1090 GCGACCTGTGCAAGCTTATGCAAGCTCCCTCCACCGCTTTCGTCATTTATGTCAG 1149  
DB 230 TGGAACTGCGGAGTGTGCGCGCACTGAAGACCGCATCATCGCCACCGTGAATCGCGAG 289  
QY 1150 GCGCTTGGCGGCTGCGCGCGCTGCGCATGTCGCCCGCATCTGATCTAGCGCGCGAAAC 1209  
DB 290 GCGGCTTCCGCGCGCGCGCGCTGCGCATGTCGTGTCGACATGTAACATGCTGCAATAC 349  
QY 1210 GCGTGGCTTCCGCGATTCGCGACCAAGGCGCGCTGCGCATCTTCTTCGCGCAACCA 1269  
DB 350 TCCACCTATTCGATGATTCGCGCGGAGGCTGCGCTCCATCTCTGGAAGACCGCGAG 409  
QY 1270 CACGCGCGGGAATCATATGAGGACAGGCGCTGCGAGCGCGCATTTTAAGCCAAAGG 1329  
DB 410 AAGGCGCGGAAAGCGCGCGCATGCGCATCAACCGCGAGCGCGTGAAGAGCGCTGGCG 469  
QY 1330 CTTATCGACGGGATCGTCGCGCAAC 1355  
DB 470 ATCGTCGACAAAGTCATGACGAGAC 495

RESULT 10  
LOCUS B1725661 539 bp mRNA linear EST 19-SEP-2001  
DEFINITION 1031080C09.y1 C. reinhardtii CC-1690. Stress II (normalised).  
ACCESSION B1725661  
VERSION B1725661.1 GI:15701356  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadaceae; Chlamydomonas.  
1 (bases 1 to 539)  
Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J. P., Shrager, J., Sillflow, C. and Stern, D.  
Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vacuolar Plants. Project: 1031  
Unpublished  
CONTACT: Charles Hauser  
DCMB Box 91000  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.  
Location/Qualifiers  
1..539  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"

FEATURES  
SOURCE

/db xref="taxon:3055"  
/clone lib="C. reinhardtii CC-1690. Stress II (normalised)  
Lambda Zap II"  
/note="vector: pBluescript II SK-; Site 1: EcoRI; Site 2:  
XhoI; Stress condition II library, constructed by John  
Davies and Jeffrey McDermott, combines cDNAs from CC-1690  
cells grown to mid-log phase in TAP (NH4+ - containing)  
and shifted to TAP - NO3- (24hrs); H2 production  
conditions (0, 12hr, 24hr) see Meis et al., (2000) Plant  
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +  
sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).  
PolyA mRNA was purified from each sample, pooled and cDNA  
synthesized. The cDNA was directionally cloned into lambda  
Zap II (Stratagene) in the EcoRI (5') and XhoI (3').  
sites. pBluescript II SK- plasmids were excised from the  
lambda Zap clones by superinfection with Exsist  
(Stratagene) phage. The library was normalized using  
method 4 described in Bonaldo et al., (1996) Genome  
Research 6: 791-806."  
BASE COUNT 95 a 185 c 175 g 84 t  
ORIGIN

Query Match  
Best Local Similarity 51.0%; Pred. No. 3e-10; Length 539;  
Matches 210; Conservative 0; Mismatches 201; Indels 1; Gaps 1;

QY 942 CACGCTTGGCGCGAGAGCTGCTTGGCGGTGTCGATTCGTCGCGCGAGCT 1001  
DB 56 CATGCCCGGCGCAAGCGCTACCGGAGCGGTGCTTATGCGCGCAAGATT 115  
QY 1002 AAACCTGCGATCGGTTCATCATGACACCTCGCGCGCGAATGTCGCGCGGCTGA 1061  
DB 116 TGTCTGCCATCATCATCTTCTGTGACACCGCGGAGCTTATGCCGCAACCGCGGA 175  
QY 1062 GAGAGCTGCGCATTCGCAAGCTCGCGCACTTGTTCAGCTTATGACGCTGCCCT 1121  
DB 176 GAGAGCTGCGCGAGGCGAGCGCATTCGCGTGAACCTGCGTGAATGTTCCGCTGCGT 235  
QY 1122 CCCCACCGTTTGGCGATTTATGTCAGGCGGTTGCGCGTGGCGCTGCGCATCTGCC 1181  
DB 236 GCCCATCATCTTGTGTCATTCGCGAGGCGGCGCTGCGCGCGCGCTGCGCATTTGGCTG 295  
QY 1182 GCGCATCTGCTTACGCGCGCGGAAAGCGGTGCTGCCGATTCGACCAAGAGGCGC 1241  
DB 296 CGCGAAGCGCAACCTGATCATGAGAGCGCGCTCTACTACGTGGCTGCCGAGCGCTG 355  
QY 1242 CTCGGCATCTCTTCCGCGA-CACCAACCAAGCGCGCGAATCATATAGCGACAAAGCG 1300  
DB 356 CGCGCGCATCTCTGCGGAGAGCGCGCTGCGCGCGCGGAGCGCATGAGCGCTGCGCG 415  
QY 1301 TCGAGCGCGACGCACTTTAAGCCAAAGGCTTATGACGCGGATCGTCGCGA 1352  
DB 416 TCACCTCGCGCGAGCTGTGAAGTTCCGCGTCATGACCAATCGTGC CGA 467

RESULT 11  
LOCUS AG247247 652 bp DNA linear GSS 13-DEC-2002  
DEFINITION Locus Japonicus DNA, clone: Ljt20k03\_not, genomic survey sequence.  
ACCESSION AG247247  
VERSION AG247247.1 GI:26647012  
KEYWORDS GSS.  
SOURCE Locus japonicus  
ORGANISM Locus japonicus  
Locus japonicus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Lotese;  
Locus.  
1  
Sato, S., Nakamura, Y. and Tabata, S.  
Locus japonicus TAC End sequences  
Published Only in Database (2002)  
2 (bases 1 to 652)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE

**AUTHORS** Sato, S.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (20-NOV-2002) Shusei Sato, Kazusa DNA Research Institute,  
 The First Laboratory for Plant Gene Research; 2-6-7  
 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan  
 (E-mail: sato@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/,  
 Tel: 81-438-52-3335 (ex. 2335), Fax: 81-438-52-3934)

# FEATURES

**Source**  
 Location/Qualifiers  
 1. .652  
 /organism="Lotus japonicus"  
 /mol\_type="genomic DNA"  
 /strain="Miyakojima MG-20"  
 /db\_xref="taxon:34305"  
 /clone="Lj120K03\_not"  
 /clone\_1db="genomic TAC library"  
 /note="VECTOR: pYLAC7"  
**BASE COUNT** 198 a 107 c 129 g 218 t  
**ORIGIN**

Query Match 5.3%; Score 77.6; DB 29; Length 652;  
 Best Local Similarity 50.0%; Pred. No. 5.2e-10;  
 Matches 222; Conservative 0; Mismatches 219; Indels 3; Gaps 1;

212 AGGCGATTCGGTACCGGCTATTTTGTCCGATTTTCTCCTCGGCGGCTTTGGGCA 271  
 72 ATGGATATTCCTGAGCAATGGGATTTATGATTTGAGTTATGGAGTATGATGGAT 131  
 272 CGGTGCGCTCGGTGCGATCATGAAGCGGATTCACCGGCGCAGAGCTGAACCTCCAC 331  
 132 CCGTAGTAGTGAGAAATCAACCGGTTGTCGATATGATACCAACCACTTTATCTTC 191  
 332 TCGTGGTCTCCCTGCTTCGGGTGTGGCGCATGCGAAGACAAATGACCTTTTCA 391  
 192 TTATGTAGATGATGATCTGAGAGACGATGACAAAGAGATTTGATTTGATGC 251  
 392 TGATGCTGCATTAACCGGCGTGTGACGCTCAACCGGAGCGCATTTGCGTTCTCG 451  
 252 AAATGCGTAAATATCTTCTGCTTATATGATTAATCACTAAATTAAGTATTTATG 311  
 452 TGTATTTGCGCA--TCCACGATGGTGGCGCATGCGCTCGTGGGCTTATCGGC 508  
 312 TATCATATCTTAATCTCCCACTAGTGGGTGAACAGCTTAGTTTGGCAATGTTGGGG 371  
 509 ATTCACCTTTTGGGAACCGCGCGCATAGTTTCTGGGCTCGCGTGGTGAAGT 568  
 372 ATATATATTGCGCAACCAATGCTTACATGCAATTTGCGGTTAAAGATTAATTGAC 431  
 569 TAACCACTGGGCGATCGCTTCAGACGGTGTGACAGCGGAGATTTGTGAAGCTG 628  
 432 AAACCTTGAATAGGCAAGTCCCGAAGTTTCAAGCGCGCAGAAATTTATTCATAGG 491  
 629 GTGTGATGATGAATGTTGTTCG 652  
 492 GCTTGTGATTCATCGTACAC 515

**RESULT 12**  
 CB977119/c 617 bp mRNA linear EST 01-MAY-2003  
**LOCUS** CB977119/c  
**DEFINITION** CAB40003 Iva\_Ra\_D11 Cabernet Sauvignon Berry - CAB4 Vitis vinifera  
 CDNA clone CAB40003\_Iva\_Ra\_D11 3', mRNA sequence.  
**ACCESSION** CB977119  
**VERSION** CB977119.1 GI:30300405  
**KEYWORDS** EST.  
**SOURCE** Vitis vinifera  
**ORGANISM** Vitis vinifera  
**REFERENCE** Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
 ; Vitaceae; Vitis;  
 1 (bases 1 to 617)  
**AUTHORS** Goes da Silva, F., Iandolo, A., Lim, H., Baek, J., Jones, K. and Cook  
 'D'.  
**TITLE** Expressed sequence tags from Vitis vinifera 'Cabernet sauvignon'

**JOURNAL** unpublished  
**COMMENT** Contact: Douglas Cook, PhD  
 CAES Genome Facility  
 UC Davis, Plant Pathology  
 One Shields Ave, Davis, CA 95616, USA  
 Tel: 530 754 6561  
 Fax: 530 754 6617  
 Email: drccook@ucdavis.edu

# FEATURES

**Source**

Location/Qualifiers  
 1. .617  
 /organism="Vitis vinifera"  
 /mol\_type="mRNA"  
 /cultiivar="Cabernet Sauvignon"  
 /db\_xref="taxon:29760"  
 /clone="CAB40003\_Iva\_Ra\_D11"  
 /sex="Hermaphrodite"  
 /dev stage="Berry on stage II, 9 mm"  
 /lab host="DH5alpha"  
 /note="Organ: Berry; Vector: pDNR; Site\_1: Sfil; Site\_2:  
 Sfil; CAB4 is a CDNA library of Vitis vinifera cv.  
 'Cabernet Sauvignon' Clone 8 berries. Sampled berries were  
 collected from field-grown vines during stage II of berry  
 growth (berries were green and hard) at approximately 60  
 days after full bloom. The average berry size was 9  
 millimeters. Sampled vines were located at the University  
 of California, Davis, Experimental Vineyard. cDNAs were  
 made by oligo-dT priming and directionally cloned. 5' and  
 3' adaptors were used in cloning as follows:  
 5'-AGCAGGATATCAACGACAGTGGCCATTCAGCGCGG-3' and  
 5'-ATTCTAGAGCGGCGGCGGCGGATG-3'. Library was  
 constructed using the Clontech Creator SMART kit and  
 size-selected to contain the 0.5-3 kb size fraction."

**BASE COUNT** 191 a 129 c 94 g 203 t  
**ORIGIN**

Query Match 5.1%; Score 75; DB 14; Length 617;  
 Best Local Similarity 47.8%; Pred. No. 2.6e-09;  
 Matches 250; Conservative 0; Mismatches 270; Indels 3; Gaps 1;

133 GGCTATGACAGACCTTGAGCGGCTGAGCAAGCGCAATGCGATCGTAAATT 192  
 616 GGTATTAAGAGCGTATGATTTTATCAAGAAAGACAGATTAATCTGAGCGCTGCA 557  
 193 ACTGGAAGACCGCGGAGCGATCCGCTAGCGCTTATTTGTCGATTTTCTTC 252  
 556 ACAAGCACAGTCAACTAAACGGTATTCGGTACAAATGGGTTATGATTTTCACTT 497  
 253 CTGCGCGTCTTTGGGACGCGTGGCGCATCAATGAAGCGGATTCACCGGCGC 312  
 496 ATGGGGGTATATGGAATCGTATGATGAGAAATTTACCGTTGATGAAATATGCC 437  
 313 ACAGAGCTGAACCTCCACTGCTGCTCCCTGCTTCGGTGTGGCGCATCAGAA 372  
 436 ACCAATGAATTTCACTCTTATTTAGTGTCTTCGAGAGACAGCATGACAGAA 377  
 373 GACATGAGCTTTTGCATGATGATGTCATTAACCGGCGTGTGACGCGTCAACCGGAG 432  
 376 GAAAGTTGAGCTGATGCAATGCGTAAATATCTTCGCTTTATATGATTAATCA 317  
 433 GCGCATTTGCGGCTCTGCTGATTTTGGCA--ATCCACGATGGGTGGCGCATGCGC 489  
 316 AATAAAGATTAATTTATGATCAATCGTCAATCTCTACTAGTGGGGGTGACAGCA 257  
 490 TCGGGGCTCATCTGGCATCTCACTTTTGGGAACCGGCGCGCATATGTTCTCG 549  
 256 AGTTTGTATGTTGGGAGATATCATTAATGCCAACCATTTCTCATATTGCGG 197  
 550 GGTCTCGCGGTGGGATTAACACATGGGCGATCGGCTTCAGACGGTGTGCACAGCGG 609  
 196 GTTAAAGGTATTAACCAATTTGAAGAAAGACAGTACTGTAAGTTTCAAGCGGCT 137



Db		122	CGGGTAAAGGCTATTGTACACCAATTTGAAAAAAGCCGTGCCTTGAGSTTCACACAGAG	63
Qy		608	CGGAGAATTTGGTGA AAACTGCTGTGATGTATGTAATTTGTGTGCCACTCCAAATTGCGTG	667
Db		62	CTGAATCTTTATACGTAAAGGGCTTATTTGATGATCATTAATGAAATTCACACNAATGAGACT	3
Qy		668	CA 669	
Db		2	CA 1	
RESULT_15				
CBSJ46970				
LOCUS				
DEFINITION				
viñifera cDNA clone CAB2SG0002_ivaf_C02 5', mRNA sequence.				
CB346970				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
Vitis vinifera				
Vitis vinifera				
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids				
; Vitaceae; Vitis				
1 (bases 1 to 606)				
Goes da Silva,F., Iandolino,A., Lim,H., Baek,J., Jones,K. and Cook				
,D.				
Expressed sequence tags from cabernet sauvignon berries at various				
developmental stages				
Unpublished				
On Mar 14, 2003 this sequence version replaced gi:28967937.				
Contact: Douglas Cook, PhD				
CAB2SG is a CDNA library of Cabernet Sauvignon Clon				
UC Davis, Plant Pathology				
One Shields Ave, Davis, CA 95616, USA				
Tel: 530 754 6561				
Fax: 530 754 6617				
Email: drcook@ucdavis.edu				
Seq primer: ACGGTACCGACATATGCC.				
Location/Qualifiers				
1..606				
/organism="Vitis vinifera"				
/mol_type="mRNA"				
/cultivar="Cabernet Sauvignon"				
/db_xref="taxon:29760"				
/clone="CAB2SG0002_ivaf_C02"				
/sex="Hermaphrodite"				
/dev_stage="veraison"				
/lab_host="PHSalpha"				
/clone_lib="Cabernet Sauvignon Berry - CAB2SG"				
/note="Organ: Berry; Vector: pDNR; Site_1: Still; Site_2:				
Still; CAB2SG is a CDNA library of Cabernet Sauvignon Clon				
8 berries. Samples were collected at veraison (transitio				
n between stage II and stage III of berry growth) from				
field-grown vines 60 days after full bloom. Berries were				
still green but soft. Sampled vines were located at the				
University of California, Davis, experimental vineyard.				
cDNAs were made by oligo-dT priming and directionally				
cloned. 5' and 3' adaptors were used in cloning as follows				
5'-AAGCACTGCTATCAACGACGAGTGGCATTTACGGCCGG-3' and				
5'-ATTCTTAGAGCCGAGGCGGCGCACATG-dt(30)-NN-3'. Library was				
constructed using the Clontech Creator SMART kit and				
size-selected to contain the 0.5-3 kb size fraction."				
BASE COUNT				
ORIGIN				
185 a				
102 c				
130 g				
.188 t				
1 others				
Query Match				
Best Local Similarity				
Matches				
248;				
Conservative				
5.0%;				
Score				
73.6;				
DB				
14;				
Length				
606;				
Pred.				
No.6.2e-09;				
Mismatch				
269;				
Indels				
3;				
Gaps				
1				
136 TTGTGAGAACCTTGGAGCGGCTGACAACAGGCGCAATGGATGGATCGGTAATTA				
CTACT				
195				

Db 2 TATTAAGAGCGATTGATTTTATCAAGAAAGACAGATTAACTGAGCGCTTCAACA 61

Qy 196 GGAAGAGGACCGGTGAGGGCATTCGGGTACCGCTTATTTTGTCCGATTTTTCCTC 255

Db 62 GGACAGGGTCACTAAACGGATATTCGGGTACCAATGGGGTTATGATTTCACTTATG 121

Qy 256 GCGGTTCTTTTGGGACGGTCCGTCGGTGGGCATCATGAAGCGATTCAACCGGCCA 315

Db 122 GGGGGTATGATGGATTCGGTATGATGAGAAATTAACCGTTTGATCGAATATGCCACC 181

Qy 316 GAGCTAAACTCCCACTGTGTCTCCCTGCTTCGGTGTGGGGCCATGACGGAAGAC 375

Db 182 AATGAATTTCTACCTCTTAATTCATGTGTCTTCGGAGAGACGATATGCAAGAAAGA 241

Qy 376 AATCAGACTTTTGTATGATGTGTTCATTAACCGGGGTGACGCTCAACCGGAGCG 435

Db 242 AGTTTACGCTTATGAATGGCTAAATATCTTCGGTTATATGATTAATCAATCAAT 301

Qy 436 CATTTGCCGTTCTGTGTATTTGGCA--ATCCAGATAGGTGGCGCCATGCGCTCG 492

Db 302 AAAAGGTATTTTATGTATCAATCCTTACATCTCTTACTGTGTGGGTGACAGCAAGT 361

Qy 493 TGGGGTCACTGGGGATCTCACTTTTGGGGAACCGGCGGCGAGATAGGTTTCTGGGT 552

Db 362 TTTGGTATGTTGGAGATATCATTTATGCGCAACCAATTCCTCATTTGAGGAGT 421

Qy 553 CTTGCGGTGTGAGTTAAACACTGGGACATGGCTTCAGACGGTGTGCACAGGCGGAG 612

Db 422 AAAAGGTATTTGAACAACATTGAAAAAGCAGTACTGTAAGGTTCACAAGCGGCTGAA 481

Qy 613 AATTTGGTAAACGTGTGTATGATTAAGAAATGTGTGCG 652

Db 482 TATTTATTCATTAAGGCTTATTCATCTCAATCGTACAC 521

Search completed: November 13, 2003, 02:26:32  
Job time : 3245.57 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2003, 16:57:27 ; Search time 91.3457 Seconds  
(without alignments)  
7117.550 Million cell updates/sec

Title: US-10-024-370-2

Perfect score: 1473  
Sequence: 1 GTGGAAGCGTTTCGAGC.....gattgagcgtttagcgag 1473

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents NA:\*\*

1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*\*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*\*  
3: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*\*  
4: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*\*  
5: /cgn2\_6/prodata/2/ina/PCTUS\_COMB.seq:\*\*  
6: /cgn2\_6/prodata/2/ina/Backfill1.seq:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the change being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1473	100.0	1473	US-09-362-899-2	Sequence 2, App11
2	1473	100.0	2123	US-09-362-899-1	Sequence 1, App11
3	312.6	21.2	4403765	US-09-103-840A-2	Sequence 2, App11
4	312.6	21.2	4411529	US-09-103-840A-1	Sequence 1, App11
5	142	9.6	36063	US-08-311-731A-140	Sequence 140, App
6	116.4	7.9	1020	US-09-252-991A-12180	Sequence 12180, A
7	116.4	7.9	1338	US-09-252-991A-12415	Sequence 12415, A
8	108.6	7.4	23673	US-09-173-816-1	Sequence 1, App11
9	104.2	7.1	870	US-09-107-532A-170	Sequence 170, App
10	97.2	6.6	444	US-09-252-991A-12365	Sequence 12365, A
11	96.2	6.5	19702	US-08-961-527-7	Sequence 7, App11
12	96.2	6.5	1830121	US-09-557-884-1	Sequence 1, App11
13	96.2	6.5	1830121	US-09-643-990A-1	Sequence 1, App11
14	93.2	6.3	2561	US-09-347-878-25	Sequence 25, App1
15	87.4	5.9	954	US-08-973-275-2	Sequence 2, App11
16	86.4	5.9	903	US-09-328-352-1114	Sequence 1114, Ap
17	85.6	5.8	1230025	US-09-198-452A-1	Sequence 1, App11
18	84.8	5.8	954	US-08-973-275-3	Sequence 3, App11
19	84	5.7	10303	US-09-634-238-410	Sequence 410, App
20	79.6	5.4	999	US-09-252-991A-4651	Sequence 4651, App
21	79.6	5.4	1203	US-09-252-991A-4229	Sequence 4229, Ap
22	77.8	5.3	456	US-09-252-991A-12319	Sequence 12319, A
23	75.2	5.1	882	US-09-134-001C-966	Sequence 966, App
24	72.8	4.9	2255	US-08-714-918-105	Sequence 105, App
25	72.8	4.9	2255	US-09-265-315-105	Sequence 105, App
26	72.8	4.9	2255	US-09-265-315-105	Sequence 105, App
27	72.8	4.9	2255	US-09-266-417-105	Sequence 105, App

28	54.6	3.7	3183	1	US-08-849-212-3	Sequence 3, App11
c 29	51	3.5	2894	1	US-08-278-091-1	Sequence 1, App11
c 30	51	3.5	2894	1	US-08-483-859-1	Sequence 1, App11
c 31	51	3.5	2894	1	US-08-472-173-1	Sequence 1, App11
c 32	51	3.5	2894	2	US-08-487-167-1	Sequence 1, App11
c 33	51	3.5	2894	2	US-08-482-816-1	Sequence 1, App11
c 34	51	3.5	2894	2	US-08-296-149-1	Sequence 1, App11
c 35	51	3.5	2894	2	US-08-801-199-1	Sequence 1, App11
c 36	51	3.5	2894	2	US-08-615-271-1	Sequence 1, App11
c 37	51	3.5	2894	3	US-09-074-660-1	Sequence 1, App11
c 38	51	3.5	2894	3	US-09-074-659-1	Sequence 1, App11
c 39	51	3.5	2894	3	US-09-106-465A-1	Sequence 1, App11
c 40	51	3.5	2894	3	US-09-106-465A-1	Sequence 1, App11
c 41	51	3.5	2894	3	US-09-106-465A-1	Sequence 1, App11
c 42	49.8	3.4	789	4	US-09-107-532A-1183	Sequence 1183, Ap
c 43	48.2	3.3	984	1	US-08-611-107-11	Sequence 11, App1
c 44	48.2	3.3	984	2	US-08-422-560A-11	Sequence 11, App1
c 45	48.2	3.3	984	3	US-08-468-793-11	Sequence 11, App1

## ALIGNMENTS

```
RESULT 1
US-09-362-899-2
; Sequence 2, Application US/09362899
; Patent No. 6361986
; GENERAL INFORMATION:
; APPLICANT: Degussa-Hls AG
; APPLICANT: Forschungszentrum-Jlich GmbH
; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY FERMENTATION AND
; CURRENT APPLICATION NUMBER: US/09362,899
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: DE 19924365.4
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1473
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1473)
; OTHER INFORMATION: accda
US-09-362-899-2

Query Match      100.0%; Score 1473; DB 4; Length 1473;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GTGGAAGCGTTTCCGACTATGTGTGGGCGATGGAACACATTCAGATTGACGCTC 60
        1 GTGGAAGCGTTTCCGACTATGTGTGGGCGATGGAACACATTCAGATTGACGCTC 60
DB      1 GTGGAAGCGTTTCCGACTATGTGTGGGCGATGGAACACATTCAGATTGACGCTC 60
QY      61 ATAGACTGCTGTTTGACCTGACGCTTCATTTCTTGGAATGAAATCCCAATATGAC 120
        61 ATAGACTGCTGTTTGACCTGACGCTTCATTTCTTGGAATGAAATCCCAATATGAC 120
DB      61 ATAGACTGCTGTTTGACCTGACGCTTCATTTCTTGGAATGAAATCCCAATATGAC 120
QY      121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGCCAAATGCGAT 180
        121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGCCAAATGCGAT 180
DB      121 AACCTCAATCAAGGCTATGACAGACCTTGGAGCGGCTCGAAGCAAGCCAAATGCGAT 180
QY      181 GAATCGTAATTAATCTGGAAGAGGCAACCGTGAAGGCGATTCGGTAGCCGTTATTTGTC 240
        181 GAATCGTAATTAATCTGGAAGAGGCAACCGTGAAGGCGATTCGGTAGCCGTTATTTGTC 240
DB      181 GAATCGTAATTAATCTGGAAGAGGCAACCGTGAAGGCGATTCGGTAGCCGTTATTTGTC 240
QY      241 GATTTTCTTCTCTCGGCGGCTTTTGGGACAGGTCGTCGTCGTCGATATGAAGCG 300
        241 GATTTTCTTCTCTCGGCGGCTTTTGGGACAGGTCGTCGTCGTCGATATGAAGCG 300
DB      241 GATTTTCTTCTCTCGGCGGCTTTTGGGACAGGTCGTCGTCGTCGATATGAAGCG 300
QY      301 ATTACCGCGCGACAGAGCTGAAACTCCCACTGCTGCTCCCTGCTTCGCGTGTGCG 360
        301 ATTACCGCGCGACAGAGCTGAAACTCCCACTGCTGCTCCCTGCTTCGCGTGTGCG 360
DB      301 ATTACCGCGCGACAGAGCTGAAACTCCCACTGCTGCTCCCTGCTTCGCGTGTGCG 360
```

301 ATTGACCGCGGCACAGAGCTGAAACTCCACTGCTGCTCTCCCTGCTTCCGGTGTGCG 360  
361 CGCATGACGAGAAAGCAATCGAGCTTTTGTGATGATGATGTCATTAACCGGCGTGTGAG 420  
361 CGCATGACGAGAAAGCAATCGAGCTTTTGTGATGATGATGTCATTAACCGGCGTGTGAG 420  
421 CGTCAACCGCGAGGCGCATTTTCCGCTCTGCTGATTTTGGCGCAATCCAGATGATGCGC 480  
421 CGTCAACCGCGAGGCGCATTTTCCGCTCTGCTGATTTTGGCGCAATCCAGATGATGCGC 480  
481 GCGATGCGCTGTGGGGGTTTCACTTGGGATCTCATTTTGGGAAACCGGCGCGAGATA 540  
481 GCGATGCGCTGTGGGGGTTTCACTTGGGATCTCATTTTGGGAAACCGGCGCGAGATA 540  
541 GGTTCCTGAGGTCTCTGCGTGTGTGAGTTAAACCACTGAGCATGCGCTTCCAGACGCTGTG 600  
541 GGTTCCTGAGGTCTCTGCGTGTGTGAGTTAAACCACTGAGCATGCGCTTCCAGACGCTGTG 600  
601 CAGCAGCGGAGAAATTTGTGTGAAAATGTGTGATTTGATGAAATTTGTGTGTGCGCACTCGAA 660  
601 CAGCAGCGGAGAAATTTGTGTGAAAATGTGTGATTTGATGAAATTTGTGTGTGCGCACTCGAA 660  
661 TTGCGTGCAGCGGTGCGCAAAAACCTCAAGTTTATTCAGCCGCTGAGAGCAACGATGCT 720  
661 TTGCGTGCAGCGGTGCGCAAAAACCTCAAGTTTATTCAGCCGCTGAGAGCAACGATGCT 720  
721 TTTTCTCCAACTCTCTGCGTGTGCACTTCCGCTGATGAGAGCGATTTGCGCTTCTGT 780  
721 TTTTCTCCAACTCTCTGCGTGTGCACTTCCGCTGATGAGAGCGATTTGCGCTTCTGT 780  
781 GACCCGCGAGAGGCGCTGAGATCGGGAGATTAATGAGAAAGTTGGGGGAGACGCTGTCAAG 840  
781 GACCCGCGAGAGGCGCTGAGATCGGGAGATTAATGAGAAAGTTGGGGGAGACGCTGTCAAG 840  
841 CTCTCTGCTGCGCTGTGCTGCGCATTTGAGCCCGCTGTGCGCTGTGCGCTGTGCGCATC 900  
841 CTCTCTGCTGCGCTGTGCTGCGCATTTGAGCCCGCTGTGCGCTGTGCGCTGTGCGCATC 900  
901 GGGGGCGCGCGCTGTGCTGCTGATTTGGGAGATGCGCGCTTCAACGCTTGGCGCGAGAG 960  
901 GGGGGCGCGCGCTGTGCTGCTGATTTGGGAGATGCGCGCTTCAACGCTTGGCGCGAGAG 960  
961 CTGCGCTTTGGCGCTGTGCGCATTTTCCGCTGCGCGCGAGCTTAAACCTGCGCATGCTGTCC 1020  
961 CTGCGCTTTGGCGCTGTGCGCATTTTCCGCTGCGCGCGAGCTTAAACCTGCGCATGCTGTCC 1020  
1021 ATCATCGACACCTCCGCGCGCGCAATTTGTGCGAGCGCGCTGAGAGCTCGGATCGCAAGC 1080  
1021 ATCATCGACACCTCCGCGCGCGCAATTTGTGCGAGCGCGCTGAGAGCTCGGATCGCAAGC 1080  
1081 TCGATTGCGCGCATCTTTGTCCAAAGCTTATCGAGCTTCCCTCCCTCCACCGCTTTCGCTCAT 1140  
1081 TCGATTGCGCGCATCTTTGTCCAAAGCTTATCGAGCTTCCCTCCCTCCACCGCTTTCGCTCAT 1140  
1141 ATTGCTCAGGGCGTGTGCGCGCGCGCTGCGCATTTGCGCGCGAGCTGCGATGCTGCTAAGCG 1200  
1141 ATTGCTCAGGGCGTGTGCGCGCGCGCTGCGCATTTGCGCGCGAGCTGCGATGCTGCTAAGCG 1200  
1201 GCGGAAAACGCGTGTGCTGTCCGATTTGCGCACAGAGGGGCGCTCGGCGCATCTCTTCCGC 1260  
1201 GCGGAAAACGCGTGTGCTGTCCGATTTGCGCACAGAGGGGCGCTCGGCGCATCTCTTCCGC 1260  
1261 GACACCAACACGCGCGCGGAAATCATAGAGCAAGAGGCTGCAAGCGCATCTTTTA 1320  
1261 GACACCAACACGCGCGCGGAAATCATAGAGCAAGAGGCTGCAAGCGCATCTTTTA 1320  
1321 AGCCAAAGGCTTATGAGCGGGATGCTGCGCGGAAACGAGCATTTTGTGAAGAAATTCG 1380  
1321 AGCCAAAGGCTTATGAGCGGGATGCTGCGCGGAAACGAGCATTTTGTGAAGAAATTCG 1380  
1381 GGCACAAATCAGCAAGCGCTCTCCGATTTGATTAACAATCCGAGAGGGCGGAGCGGAC 1440  
1381 GGCACAAATCAGCAAGCGCTCTCCGATTTGATTAACAATCCGAGAGGGCGGAGCGGAC 1440

1441 AGTCGCTTCAACAGATTGAGCGTTTACGCGAG 1473  
1441 AGTCGCTTCAACAGATTGAGCGTTTACGCGAG 1473  
RESULT 2  
US-09-362-899-1  
Sequence 1, Application US/09362899  
Patent No. 6361986  
GENERAL INFORMATION:  
APPLICANT: Degussa-Hls AG  
APPLICANT: Forschungszentrum-Jlich GmbH  
TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY FERMENTATION AND  
CURRENT FILING DATE: 1999-07-29  
EARLIER APPLICATION NUMBER: DE 19924365.4  
EARLIER FILING DATE: 1999-05-27  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 1  
LENGTH: 2123  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
FEATURE:  
NAME/KEY: gene  
LOCATION: (508)..(1380)  
OTHER INFORMATION: accda  
US-09-362-899-1  
Query Match 100.0%; Score 1473; DB 4; Length 2123;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 GTGAGAGAGCGTTTCCGACTGATGTGTG3GCGATGGAACAACATTGACGCTC 60  
508 GTGAGAGAGCGTTTCCGACTGATGTGTG3GCGATGGAACAACATTGACGCTC 567  
61 ATAACCTCGGTTTGGACCTTGACAGCTTCACTTTCTTGGAATGAATCCCAATATGAC 120  
568 ATAACCTCGGTTTGGACCTTGACAGCTTCACTTTCTTGGAATGAATCCCAATATGAC 627  
121 AACCTCATCAAGCTTGTGACAGACCTTGAAGCGGCTCGAAGCAAGCCAAATGCGAT 180  
628 AACCTCATCAAGCTTGTGACAGACCTTGAAGCGGCTCGAAGCAAGCCAAATGCGAT 687  
181 GAATCGGTAATTACTGAGAAAGGCACTGTGAGGCAATTCGCTAGCCGTTATTTGTCC 240  
688 GAATCGGTAATTACTGAGAAAGGCACTGTGAGGCAATTCGCTAGCCGTTATTTGTCC 747  
241 GATTTTCTTCTTCCGCGCTTCTTGGGACGCTTCCGCTGCGCATGAGAGCG 300  
748 GATTTTCTTCTTCCGCGCTTCTTGGGACGCTTCCGCTGCGCATGAGAGCG 807  
301 ATTACCGCGCAGAGAGCTGAAACTCCACTGCTGTCTCCCTGCTTCCGCTGTGCG 360  
808 ATTACCGCGCAGAGAGCTGAAACTCCACTGCTGTCTCCCTGCTTCCGCTGTGCG 867  
361 CGCATGACGAGAAAGCAATCGAGCTTTTGTGATGATGATGTCATTAACCGGCGTGTGAG 420  
868 CGCATGACGAGAAAGCAATCGAGCTTTTGTGATGATGATGTCATTAACCGGCGTGTGAG 927  
421 CGTCAACCGCGAGGCGCATTTTCCGCTCTGCTGATTTTGGCGCAATCCAGATGATGCGC 480  
928 CGTCAACCGCGAGGCGCATTTTCCGCTCTGCTGATTTTGGCGCAATCCAGATGATGCGC 987  
481 GCGATGCGCTGTGGGGGTTTCACTTGGGATCTCATTTTGGGAAACCGGCGCGAGATA 540  
988 GCGATGCGCTGTGGGGGTTTCACTTGGGATCTCATTTTGGGAAACCGGCGCGAGATA 1047  
541 GGTTCCTGAGGTCTCTGCGTGTGTGAGTTAAACCACTGAGCATGCGCTTCCAGACGCTGTG 600  
1048 GGTTCCTGAGGTCTCTGCGTGTGTGAGTTAAACCACTGAGCATGCGCTTCCAGACGCTGTG 1107



QY	601	CAGCAGGCGGAGAAATTGGTGAAGAACTGGTGTGATTGAAGAAATTGTGTGTGCGACATCCAA	660
Db	1108	CAGCAGGCGGAGAAATTGGTGAAGAACTGGTGTGATTGAAGAAATTGTGTGTGCGACATCCAA	1167
QY	661	TTGCGTGCAGCGGTGGCAAAAACCTCAAGGTTATTACGCGGTGTAGAGGCAACGATTCGT	720
Db	1168	TTGCGTGCAGCGGTGGCAAAAACCTCAAGGTTATTACGCGGTGTAGAGGCAACGATTCGT	1222
QY	721	TTTTTCCAAACAACCTCTGGCGGTGCACCTTCCGGTGTATGGAAGCGATTTCGCGGTTCTGCT	780
Db	1228	TTTTTCCAAACAACCTCTGGCGGTGCACCTTCCGGTGTATGGAAGCGCATTCGCGGTTCTGCT	1287
QY	781	GACCCGCAAGAGGCTGTGAATCGGGAGATTTATGAAAACGTTGGGGCAGACGTGTCCAG	840
Db	1288	GACCCGCAAGAGGCTGTGAATCGGGAGATTTATGAAAACGTTGGGGCAGACGTGTCCAG	1347
QY	841	CTTTTCGTGTGCGCGTGTGCGGCATTTAGCCCGCGCTGTGCGCGTTTCCTTGCGCGCATC	900
Db	1348	CTTTTCGTGTGCGCGTGTGCGGCATTTAGCCCGCGCTGTGCGCGTTTCCTTGCGCGCATC	1407
QY	901	GCGGCGCCGACCGCTGTGTGCTGATTTAGGACAGATTCGCGCGTTTACGCTTGGGCGCGAGAG	960
Db	1408	GCGGCGCCGACCGCTGTGTGCTGATTTAGGACAGATTCGCGCGTTTACGCTTGGGCGCGAGAG	1467
QY	961	CTGCGTTTTCGCGCGTGTGTGACATTTTCGCTGTGCGCGCAGACTTAAACTGTGCCATTCGTGTC	1020
Db	1468	CTGCGTTTTCGCGCGTGTGTGACATTTTCGCTGTGCGCGCAGACTTAAACTGTGCCATTCGTGTC	1527
QY	1021	ATCATCGACACCTCGCGGCGCGCAATTTGTGTGCGAGGGCGGTGAGAGACTGTGGCATTCGCAAG	1080
Db	1528	ATCATCGACACCTCGCGGCGCGCAATTTGTGTGCGAGGGCGGTGAGAGACTGTGGCATTCGCAAG	1587
QY	1081	TCGATTGGCGCGACCTTGTTCCAAGCTTTACGACGCTCCCTTCCACACGTTTCGTGTCATT	1140
Db	1588	TCGATTGGCGCGACCTTGTTCCAAGCTTTACGACGCTCCCTTCCACACGTTTCGTGTCATT	1647
QY	1141	ATTGTGAGGGCGCTTGTGGCGGTGTGCGCGCTGTGGCATTCGTGCTGCCCGCATTCGTGTACGCG	1200
Db	1648	ATTGTGAGGGCGGTGTGGCGGTGTGCGCGCTGTGGCATTCGTGCTGCCCGCATTCGTGTACGCG	1707
QY	1201	GCCGAAAACGGGTGTGCTGTGCGCATTTGCGACACAGAGGGCGCGCTCGGCATTCCTTTCGCG	1260
Db	1708	GCCGAAAACGGGTGTGCTGTGCGCATTTGCGACACAGAGGGCGCGCTCGGCATTCCTTTCGCG	1767
QY	1261	GACACCAACCAACGCTCGCGGAAATCATAGACGACAAGGCGTGTACAGGCGCACGCACTTTTA	1320
Db	1768	GACACCAACCAACGCTCGCGGAAATCATAGACGACAAGGCGTGTACAGGCGCACGCACTTTTA	1827
QY	1321	AGCCAAAGGGCTTATGACGGGATGTGTGCGCGAAAACCGAGCATTTTGTGGAAGAAATTTC	1380
Db	1828	AGCCAAAGGGCTTATGACGGGATGTGTGCGCGAAAACCGAGCATTTTGTGGAAGAAATTTC	1887
QY	1381	GGCACAATCAGCAACGCCCTTCTCCGAATTGATTAACAATCCGAGAGGGCGCGGACGCGAC	1440
Db	1888	GGCACAATCAGCAACGCCCTTCTCCGAATTGATTAACAATCCGAGAGGGCGCGGACGCGAC	1947
QY	1441	AGTGCCTTCAACAATTTGAAGCTTTAGCGCAG	1473
Db	1948	AGTGCCTTCAACAATTTGAAGCTTTAGCGCAG	1980

```

RESULT 3
US-09-103-840A-2/C
; Sequence 2, Application US/09103840A
; Patent No. 6294328
;
; GENERAL INFORMATION:
;
; APPLICANT: FLEISCHMAN, Robert D.
;
; APPLICANT: WHITE, Owen R.
;
; APPLICANT: FRASER, Claire M.
;
; APPLICANT: VENTER, John C.
;
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;
; TITLE OF INVENTION: TUBERCULOSIS

```

```

; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
;
; TYPE: DNA
;
; ORGANISM: Mycobacterium tuberculosis
;
; FEATURE:
;
; OTHER INFORMATION: CDC 1551
;
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
;
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

Query Match	21.2%;	Score	312.6;	DB	3;	Length	4403765;
Best Local Similarity	54.0%;	Pred. No.	1.1e-76;				
Matches	717;	Conservative	0;	Mismatches	589;	Indels	21;
						Gaps	3

Qy	65	ACTGCGTTTTGAAACCTCGACAGCTTCATTTCTTGGAAATGAAACCTCCCAATATAGAAC	124
Db	1008148	ACGCGGTGCTAGACCCGGGGAATCTTTGCTCAGCTGGGATAGCGACCCGCTGGCGTCCGG	1008089
Qy	125	TCATCAAGGCTATGACAGACCTTGGAGCGGGCTCGAACCAAGGCCAAATGCGATGAAT	184
Db	1008088	TAGCCGACTCTATAGCGCGGAGAGCTGGCCGCGCTGGGGGGGCCACCGGGCGGACGAAT	1008029
Qy	185	CGGTAAATTACTGGAAGAAAGCAACGTGAGGGGCAATTCGGTAGCCGTTATTTTGTCCGATT	244
Db	1008028	CGGTGACAGACCGGTGAGGGGACGGGTAAATCCGGCGGGCGGTGGCCGTGTGGCTGTAGAT	10079659
Qy	245	TTTCTCTTCCTCGGGGGGTCTTTGGGGGACGGGTGCGGTGCGGCGATATAGAAGCGCATTC	304
Db	1007968	TCGACTTCTCTGGGGGCGCTCGAATTTGGGGGTGGGAGGAGCGAAACGATACCGCGCCGTGC	1007909
Qy	305	ACGCGCCACACAGAGCTGAAACTCCACTGCTGGTCTCCCTGCTTCGGTGGTGCGGCA	364
Db	1007908	AGGGGGGAGACCGCGCAGCGGGCTGCGCTACTTGGCGTCAACAAGTCCGGAGAGCACCCGCA	1007849
Qy	365	TGACGAGAAACAATCGAGCTTTTGTGCAATGATGGTGTTCATPACCGCGGGCTGTGACGGCTC	424
Db	1007848	TGCAAGAAAGCACGGTGCGCTTTCTGCAAGATGGGAAGATCGCTGGCGGCATCCAGCTGC	1007789
Qy	425	ACGCGAGGGGCAATTTGGCTTCTCTGGTGAATTTGGCGAATCCAGATGGGTGGCGCCA	484
Db	1007788	ACAAACGAGGGCGCCCTGCGCTTACTGACTGTGCTAATTTGGCCATCCGACCAACGGGTGGAGTTT	1007729
Qy	485	TGGCTCGTGGGGTTACTGAGGAGATCTCACTTTTGGAGAACCGGGCGCGACAGTAGGTT	544
Db	1007728	TGCGGTGTGGGGCTGCTGGGGGATCTTCAACCGTGGCGAGCGGGGCCCTTATATCGGCT	1007669
Qy	545	TCTTGGGTCTCGCGTGGTGAATTAACAATGGGAGTGGCTGTCCAGACGGTGTGCAGC	604
Db	1007668	TTCTGGGAGCACGGGTCTATGAGATTGCTGATGGCGACCCCTTCCCATCCGGGGTCCAAA	1007609
Qy	605	AGCGGAGAAATTTGGTGAATACTGTGTGATATGAGAAATTTGTGCGCACTCCAAATGCG	664
Db	1007608	CCGCCGAGATCTACCGCGGCAATGAGGATCATCGACGGCGCTGTTCAGTACGACCGGCTAC	1007549
Qy	665	GTCGACGGGTGGGCAAAAACCTCAAGTTATTCAGCCCGTGAAGGCAACGATCGTTTTT	724
Db	1007548	GACCGATGCTGGAATCGTGGCTTAGCGGTGCTCATTCGACGCTTCCGAAACCGCTTCCGGAC	1007489
Qy	725	CTTCAACAATCTCTGGCGTGGCA---CTTCCGGTGAATGAGCGGATTTGCGCTTCTCGTG	781
Db	1007488	CGAGAGCGCCCGGCCCTGATCCCGATGTGCCACGTGGGACTGGGTGTGGATTCGCGCC	1007429
Qy	782	ACCCGAGAGGCTTGAATTCGGGAGATTATGGAACGTTGGGGCGAGACGTGTCAAGC	841
Db	1007428	GGCCGAGACCGGCGGGCGTCAAGGACATCTGAGCACAGCGGCGCACGACCGGGGTGTTG	1007369
Qy	842	TTTCTGTGCGGTGCTGGCGCATTAAGCCCGGCTGTGGCGTTGGCCCTTGGCGCGCATCG	901



Db 1007368 TGTCAGAACCGATCAAGGCGGAGCG---GGACACACGCTGCGCTGCGCCGCTTTG 1007312  
Qy 902 GGGGCGGGCCCGTGTGTGATTTGGGCGAG-----GATGCCGCTTCAAGC 946  
Db 1007311 GCGGCAACCCACGCGTGTCTCTGGCCAGAAAGGCGAGTAGGCGCGCGGAGAACACTG 1007252  
Qy 947 TTGGGCGCGAGAGCTGGGTTTTTGGCGTGTGCAATTCGTGGCGCGGAGTAAC 1006  
Db 1007251 TCGGGCCCGCTGCGTTAGCGCAAGCCGACGCGGATGGCGCTGCCCGCGAGCTGTGCC 1007192  
Qy 1007 TCGCGATGTGTCCATCATTCGACACCTCCGCGCGGAAATGTGCGAGCGGCTGAGAGC 1066  
Db 1007191 TGCCTGTGTGTGTGATTCAGCGCGCGGACCGCGCTTGTGCGCGCGAGCGAACAG 1007132  
Qy 1067 TCGGATGCGAAGCTCGAATGGCGGCGACCTTGTCAAGCTTATCGAACCTCCCTCCCA 1126  
Db 1007131 GCGGCGTGGCGCGGCGAGATCGGCAATGCGCTGTGCGGACCTGTCAGCTGATACCCGA 1007072  
Qy 1127 CCGTTTCGATCATTTATTTGATGAGGCGTGTGGCGGTGGCGCGCTGACATGCTGCCGCG 1186  
Db 1007071 CCGTGTGATCTGTGTGGCGGAGCGGCGAGCGGCGCGCGCTGCGGATGTTGCCGCG 1007012  
Qy 1187 ATCTGTCTACGCGGCGGAAAGCGGCTGTGCTGCGATTCGACAGAGGCGCTGCG 1246  
Db 1007011 ACCGCGTGTGCGCGCACTCCACGCGCTGCGCGCTTGTGCTCCGAGAGGCGAGCG 1006952  
Qy 1247 CCATCTCTTCCGCGGACCAACGCGCGGGAATGATAGAGGAGCAAGGCGTGCAG 1306  
Db 1006951 CGATGTGTTCGAGACACTGCTCATGCGCGGAGCTGCTGCCGCCAGGCGATCCGCT 1006892  
Qy 1307 CGCAGCACTTTTAAGCCAAAGGCTTATCGAGGAGTGTGCGCCGAAACGAGCACTTTG 1366  
Db 1006891 CGGCGCACTGTAAGTTCGGGATTTGTGACACATGCTGCGGAGTACCCCGAGCGCG 1006832  
Qy 1367 TTGAAGA 1373  
Db 1006831 CAGACGA 1006825

RESULT 4  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 21.2%; Score 312.6; DB 3; Length 4411529;  
Best Local Similarity 54.0%; Pred. No. 1,1e-76;  
Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;

Qy 65 ACTCGTTTGGACCTGACGCTTCATTTCTTGAATGAATCCCAATATGACAAC 124  
Db 1008147 ACGCGGTGTAGACCGGGATTTTTCGTAGCTGGATAGCAGCCGCTGGGCGTGGCGG 1008088  
Qy 125 TCAATCAAGGCTATGACAGAGACTTGGAGCGGCTTGAAGCAAGCCAAATGATGAT 184  
Db 1008087 TAGCGACTCTATGCGCGGAGCTTGGCGCGCTTGGCGGCGCACCGGCGGAGCAT 1008028

Qy 185 CGGTATTATCGAGAGGACCGGTGAGGGCAATTCGATAGCCGTTATTTGTCGATT 244  
Db 1008027 CGGTGACAGACCGGTGAGGAGACCGGTATTTCGGGCGGCGGGGTGGCGCTGTGAGT 1007968  
Qy 245 TTTCCTTCTCTGCGGCTTCTTTTGGGACCGGTGCGTGTGCGCATATGAAGCGATT 304  
Db 1007967 TCGACTTCTGTGGGCGGTCTCATATTGGGTGGACGCGGCGGAATCACCGCGCGTGG 1007908  
Qy 305 ACCGCGGACAGAGCTAAACTCCCATGTGTGTCTCCCTGCTTCCGCTGTGGCGGCA 364  
Db 1007907 AGCGGGGACCGCGGACCGGCTGCGCTACTGTGCTACCAAGCTTGGAGGACCGGCA 1007848  
Qy 365 TCGAGAAAGCAATTCGAGCTTTTGTCTGATAGTGTTCATTAACCGGCGGTGCGGCTC 424  
Db 1007847 TCGAAGAAAGCACGCTGCGCTTTCTGCAAGTGTGAAGATCGCTGCGGCCATCCAGCTGC 1007788  
Qy 425 ACCGAGGCGCATTTTGCCTGCTGTGTATTTTGGCAATCCACAGATGGTGGCGCA 484  
Db 1007787 ACAACAGGCGCGCTGCTCTTACTGTGTATTTGGCCATCCGACAGGCTGGAGTTT 1007728  
Qy 485 TGGCTTGTGGGTTTCACTTGGGATCTCACTTTTGGGAAACCGGCGGCGCATAGTT 544  
Db 1007727 TCGGCTGTGGGCTTCTGTTGGGATCTCAACGTCGCGAGCGGCGGCTTGTAGTGGCT 1007668  
Qy 545 TCTTGGGCTTCCGCGGTGTGAGTTAACACTGGGCAATGCGCTTCCAGAGGCTGCGAGC 604  
Db 1007667 TTTCTGGAACACGCGGTCTTATGATGTTGCTATGCGCACCCCTTCCATCCGCGTCCAA 1007608  
Qy 605 AGGCGAGAAATTTGTGTAAGAACTGTGTGATTAATGTAATTTGTGCGCATCTCCATTTGC 664  
Db 1007607 CCGCGAGAAATCTACGCGGCGCATGAGATTCATCAACGCGCTGTGTGCACTGACCGGCTAC 1007548  
Qy 665 GTGAGAGGTGGGAAAAACCTCAAGTTTATTCAGCGGCTGAGGCAACGATGCTTTT 724  
Db 1007547 GACGATGCTGATGCTGTGCTGATGAGGTCTATGACGCTCCGAAACCTCTCCGAC 1007488  
Qy 725 CTCACAACTCTCTGCGGTGAGCA---CTTCCGATGATGAGAGGATTTGCGTCTGTG 781  
Db 1007487 GCGAGACGCCCGCGCTGATCCGATGTGCTCACATGTGGAGCTGTGTGTGATCGCGCC 1007428  
Qy 782 ACCGCAAGAGCCTGTAATCGGGAGATTATGAAAGTTTGGGAGCGACGCTGTCAAGC 841  
Db 1007427 GCGCGAACCGGCGCGGCTGAGGAGACTACTGCGACACGCGCCACGACCGGCTGTGT 1007368  
Qy 842 TTTCTGTGGCGGCTGTGCGGATTTAGCCCGGCTGTGCGGCTTCCCTGCGGCTATG 901  
Db 1007367 TGTCAAGAACCGATCAAGCGGAAGCG---GCAACACGCTGTGCGCTGCGCTTTG 1007311  
Qy 902 GGGCGGCGCGTGTGCTGATTTGGGCGAG-----GATCGCGCTTTCAGCG 946  
Db 1007310 GCGGCAACCCACGCTGTGTCTCTGCGCAGCAAGAGGAGTAGGCGGCGGAGGAGCACTG 1007251  
Qy 947 TTGGGCGCAGAGACTCGTTTTCGCGGTGTGTGCAATTTGCTGCGCGCGGAGCTAAAC 1006  
Db 1007250 TCGGCGCGCGCTGCGTTACGGAAGCCGACGCGGAGTAGGCGCTGCGCGGAGCTGTGCC 1007191  
Qy 1007 TCGCGATGCTGTCCATATGACACCTTCGCGCGCGGAAATGTGCGAGCGGCTGAGAGC 1066  
Db 1007190 TGCCTGTGTGTGTGATTAAGACCGGCGGACCGCGGCTGTGCGCGGCGGAGCAAG 1007131  
Qy 1067 TCGGATGCGAAGCTGATTTGCGGCGGACCTTGTCAAGCTTATCGACGCTCCCTCCCA 1126  
Db 1007130 GCGGCGTGGCGGCGCAATGCGCATTTGCTTGGCGAGCTGTGATGCTGATATCCCGA 1007071  
Qy 1127 CCGTTTCGATCATTTATTTGATGAGGCGTGTGGCGGTGGCGCGCATGCTGCGCGCG 1186  
Db 1007070 CCGTGTGATCTGTCTGGGCGAGGCGAGCGGCGGCGGCGGCTGTGCGGATTTGCCGCG 1007011  
Qy 1187 ATCTGTCTACGCGGCGGAAAGCGGCTGTGCTGCGATTTGCCACAGAGGCGCTTGG 1246  
Db 1007010 ACCGCGTGTGCGCGCACTCCACGCTGCGCTGTGGCGCTTGTGCTTCCGAGAGGAGCCGCG 1006951

QY 1247 CCATCTCTTCGCGACACCAACGCGCGGAAATCATAGACGACAGCGCTGCAG 1306  
 DB 1006950 CCATCTCTTCGCGACACCAACGCGCGGAAATCATAGACGACAGCGCTGCAG 1006891  
 QY 1307 GCGACGACCTTTAAGCCAGGCGTTATGAGGAGTCGTCCGCGAACCAGCACTTTG 1366  
 DB 1006890 GCGCGACCTTACTGAAATCGGGGATGTGACACCATCGTCCGCGAGTACCCCGACGCCG 1006831  
 QY 1367 TTGAAGA 1373  
 DB 1006830 CAGACGA 1006824

RESULT 5  
 US-08-311-731A-140  
 ; Sequence 140, Application US/08311731A  
 ; Patent No. 6583266  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SMITH, DOUGLAS  
 ; APPLICANT: MO, JEN-I  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
 ; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LARPAE FOR  
 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
 ; NUMBER OF SEQUENCES: 411  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
 ; STREET: 600 ATLANTIC AVENUE  
 ; CITY: BOSTON  
 ; STATE: MASSACHUSETTS  
 ; COUNTRY: USA  
 ; ZIP: 02210  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/311,731A  
 ; FILING DATE:  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: GATES, EDWARD R.  
 ; REGISTRATION NUMBER: 31,616  
 ; REFERENCE/DOCKET NUMBER: C0044/7125  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617/720-3500  
 ; TELEFAX: 617/720-2441  
 ; INFORMATION FOR SEQ ID NO: 140:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 36063 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: circular  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHEetical: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Mycobacterium leprae  
 ; US-08-311-731A-140

Query Match 9.6%; Score 142; DB 4; Length 36063;  
 Best Local Similarity 57.1%; Pred. No. 1.3e-29;  
 Matches 296; Conservative 0; Mismatches 220; Indels 2; Gaps 2;  
 QY 178 GATGATGCTGTAATTAATGAGAGCAACCGTGAAGGCGATTCGGTATGATTTTG 237  
 DB 31425 GACGATGCTGCTGACACGAGCGGCGATTCGGGACGCGGATGCGATATGTC 31484  
 QY 238 TCCGATTTTCTCTCTCGGCGGCTTTTGGGACGCTCGCTGCTCGATCATGAAG 297  
 DB 31485 TCGAATTCAGCTTCTCTGCGGCGTATCGGGGTGGGACGCGGAGCGAATCACCGCC 31544  
 QY 298 GCGATTCAACCGCGACAGAGCTGAACTCCCACTGCTGTCTCCCTGCTTCCGCTGCT 357

DB 31545 GCGGTGACCG- GCGACAGCTGAGAGCGGTGCGC- GCTGCTGCGCCAAAGTTCCGATGGC 31602  
 QY 358 GCGCGATGACGAGAAAGCAATCGAGCTTTTGTATGATGCTGCTCATACCGCGCTGTG 417  
 DB 31603 ACCCGTATGACGAGAAAGCAACCGTCTGCTGCGAGATGAGATGCTGCGCGCTGTG 31662  
 QY 418 CAGCGTCAACCGGAGGCGCATTTGGCGTTCCTGCGGTATTTGGCGAATCCACGATGGGT 477  
 DB 31663 AACTCTCAAAACGCGCGGCTGTGCTTACTTAAAGGTAATGCGTAATCCAGACCGGC 31722  
 QY 478 GCGCGCATGCGCTCGTGGGCTTCAATCGGCGATCATCTTTGCGAAACCGCGCGAG 537  
 DB 31723 GGGCTCTTGGCGTGGGCGCTCGTGGCGCATGTAACCGTCCGAGCCGGGTGTTTA 31782  
 QY 538 ATAGTTCTCTGCGTCTCGGCTGAGTAACTGAGCATGCGCTTCCAGACGCT 597  
 DB 31783 ATCGCTTCTTGGGCGCTCGGCTGATGATGCTGATGCGAACTTCCCGTCCGAC 31842  
 QY 598 GTGCGACGAGGCGGAATTTGTTGAAACTGCTGATGATGATGATGATGCTGCTGCTG 657  
 DB 31843 ATCGACAGCGGAGAAATCTCAACGCGCATGGGTAATGACGCTATGCTGACGCTGAC 31902  
 QY 658 CAATTGCGTGCAGCGGTGCGCAAAACCTCAAGTTAT 695  
 DB 31903 GGACTGCAATGACGCTTGAATCGTGGCTGACGATGAT 31940

RESULT 6  
 US-09-252-991A-12180  
 ; Sequence 12180, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 12180  
 ; LENGTH: 1020  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-12180

Query Match 7.9%; Score 116.4; DB 4; Length 1020;  
 Best Local Similarity 52.0%; Pred. No. 4.2e-23;  
 Matches 261; Conservative 0; Mismatches 241; Indels 0; Gaps 0;  
 QY 179 ATGAATGGTAATTAATGAGAGCAACCGTGAAGGCGATTCGGTATGATTTTGT 238  
 DB 473 AGGACGCGCTGATGCGATGACGCGCAAGCTGCGAGGATGCGGCTGCGCGCT 532  
 QY 239 CCGATTTTCTCTCTCGGCGGCTTTTGGGACGCTGCGCTGCGTCCGATCATGAAG 298  
 DB 533 TCGAGTTCTCTTATATGCGGCTTTCGATGCGCGCATGCTGCGGAGCGCTTCTTGGCG 592  
 QY 299 CGATTCAACCGGCGACAGAGCTGAACTCCCATGCTGCTGCTTCCCTGCTTCCGCTGTG 358  
 DB 593 CAGCAACGTCGCCCTGAGAAAGCGCTGCCGCTGATCTCTTCCGCTCCGCGCGCG 652  
 QY 359 CCGCGATGACGAGAAAGCAATCGAGCTTTTGAATGATGCTGCTGATTAACCGCGCTGTC 418  
 DB 653 CCGCGATGACGAGAAAGCGCTGATCTGCTGATGAGAGGCGCAAGACCTCGCGGCTCTG 712  
 QY 419 AGCGTCAACCGGCGACGATTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 478  
 DB 713 CCGCGCTGCGGAGAAAGGATCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 772

Query 479 GCGCATGCTCTCGGGGTTTCATCTGGGCACTCTTTCGGGAACCGGGCGGAGA 538  
Db 773 GCGTTTCGCGCAGCGCTGGATCTCGGAGATGATCGGCAACCAAGGCGTGA 832  
Query 539 TAGGTTCTGAGTCTCGCGTGTGAGTTAACCACTGGGCACTCGTTCCAGACGCTG 598  
Db 833 TCGGCTTGGCGGCTCCCGGCTGATCGAGCAGACCTCGCGAGAAAGCTGCGGAGGCT 892  
Query 599 TGCAGCAGCGGAGAAATTGTGAAACTGTGTGATTGATGAAATTGTGTGCACTCC 658  
Db 893 TCCAGCGTAGGAGATTCTCTTGTAGCATGGCGCATGATGATGATGATGATGATGATG 952  
Query 659 AATTGCGTGCAGCGGTGGCAAA 680  
Db 953 AGTTGCGGCGCGCTGGCCAA 974

RESULT 7  
US-09-252-991A-12415/c  
Sequence 12415, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 12415  
LENGTH: 1338  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-12415

Query Match 7.9%; Score 116.4; DB 4; Length 1338;  
Best Local Similarity 52.0%; Pred. No. 4,7e-23;  
Matches 261; Conservative 0; Mismatches 241; Indels 0; Gaps 0;  
Query 179 ATGATTCGGTATTTCTGGAAGAAGCAACGTTGAGGCAATTCGGTATCTATTTTGT 238  
Db 677 AGGAGCGCTGATCGCCATGAGCGGCAAGCTGCAAGGATGCGGATGCGCTGCGCT 618  
Query 239 CCGATTTTCTCTCTCGCGGTTCTTTGGGACGCGTGGTGGGCAATCATGAAG 298  
Db 617 TCGAGTTCTCTTCAATGAGCGGTTGATGAGCGCATGCTGGCGAAGCGCTTCTTCCG 558  
Query 239 CGATTACCGCGCAAGAGCTGAACCTCCACTGCTGCTCTCCCTGCTTCCGCTGTG 358  
Db 557 CAGCCACGTCGCCCTGAGAGCGCTGCGCTGATCTGCTTCCGCTCCGCGGCG 498  
Query 359 CGCGATGAGAAAGCAATGAGCTTTGTATGATGTGTTCATTAACCGGGGCTGTC 418  
Db 497 CGCGATGAGAAAGCGCTGATCTGCTGATGCAATGCGCAAGACTTGGGCGTCTGG 438  
Query 419 AGCGTACCGCGAGAGCGCATTTGCGTTCTGTGTTATTTGGCGCATCCAGATGAGGTG 478  
Db 437 CGCGCTGCGCGAAGAGCATCCCGTTCTGCTGATTAACCAACCGGCTTACGCG 378  
Query 479 GCGCATGAGCTCTGAGGTTCTGAGCATCTCACTTTGCGGAACCGGCGCGAGA 538  
Db 377 GCGTTTCGCGCAGCGTGGATGCTCGGCGAGTATGATCGGCAACCAAGGCGCTGA 318  
Query 539 TAGGTTCTGAGTCTCTCGCGTGTGAGTTAACCACTGGGCACTCGCTTCCAGACGCTG 598  
Db 317 TCGGTTCTGCGCGCTCCCGCTGATCGAGCAGACGCTCGCGAAGAGCT 258  
Query 599 TGCAGCAGCGGAGAAATTGTGAAACTGTGTGATTGATGAAATTGTGTGCGCACTCC 658

Db 257 TCCAGCTAGCGACTTCTCTTGTAGATGCGCCCATGACATGATGATGATGATGATG 198  
Query 659 AATTGCGTGCAGCGGTGGCAAA 680  
Db 197 AGTTGCGGCGCGCTGGCCAA 176

RESULT 8  
US-09-773-816-1/c  
Sequence 1, Application US/09773816  
Patent No. 6340774  
GENERAL INFORMATION:  
APPLICANT: Stanford University  
TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR  
FILE REFERENCE: 28600-20210.00  
CURRENT APPLICATION NUMBER: US/09/773,816  
PRIOR FILING DATE: 2001-06-08  
PRIOR APPLICATION NUMBER: US 60/243,458  
PRIOR FILING DATE: 2000-10-25  
PRIOR APPLICATION NUMBER: US 60/179,305  
PRIOR FILING DATE: 2000-01-31  
NUMBER OF SEQ ID NOS: 1  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 23673  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(23623)  
OTHER INFORMATION: n = A,T,C or G  
US-09-773-816-1

Query Match 7.4%; Score 108.6; DB 4; Length 23673;  
Best Local Similarity 46.6%; Pred. No. 2,7e-20;  
Matches 611; Conservative 0; Mismatches 619; Indels 81; Gaps 5;  
Query 177 CGATGAATCGGTATTTACTGAGAGAAGCAGCGTGGAGGCAATTCGGTATCTATTTT 236  
Db 8617 CGAGAGGCGCTGCTCATCTGACCGGCGGAGATCCAGCGCAGCGCTGCTGCGCGC 8558  
Query 237 GTCCGATTTTCTCTCTCGCGGTTCTTTGGGACGCGTGGTGGGCAATCATGA 296  
Db 8557 GATGACTTCTGCTCTCTCGCGGACCTTGGGCGCGCGCGTGGCGCACTGATCATCT 8498  
Query 297 GCGCATTCACCGCGCAAGAGCTGAACCTCCACTGCTGATCTCCCTGCTTCCGCTG 356  
Db 8497 CGCGCGCGAAGCGCGCTCGCGCAACACTGCGCTGCTGATGATGATGATGATGATG 8438  
Query 357 TGGCGGATGAGAGAGCAATGAGCTTTGTATGATGATGATGATGATGATGATG 416  
Db 8437 CGCGCATGAGAGAGAGCGCGCTGCGCTCATGATGATGATGATGATGATGATGATG 8378  
Query 417 GCAAGCTACCGCGAGCGCATTTGCGCTTCTGATGATGATGATGATGATGATG 476  
Db 8377 CGGGAACCTGAGCAGAGCGCGGCTGATGATGATGATGATGATGATGATGATGATG 8318  
Query 477 TGGCGCATGAGCTCTGAGGATTCATCTGAGCATCTCACTTTTGGGAACCGGCGGCA 536  
Db 8317 CGGCGTGGCGCGCTCTCTTCTGAGACCTGATGATGATGATGATGATGATGATGATG 8258  
Query 537 GATAGTTTCTGAGTCTCTGCGTGTGAGTTAACCACTGGGCACTCGCTTCCAGACG 596  
Db 8257 TCTCGGCTTCCGCGCGCGCGGCTCATGCGCGAGACCATCCGAGAACTGCGCGCGA 8198  
Query 597 TGTGACAGCGCGAGAAATTGTGAAACTGTGTGATGATGATGATGATGATGATG 656  
Db 8197 GTTCCAGCGCGCAATACCTCTCTGAGAACGAGCTGATGATGATGATGATGATGATG 8138  
Query 657 CCAATTGCTGACAGCGGTG-----CAAAAACC 685

Db 8137 CGGCTGCGCCGACGAACCTGCGCCGCTGCGCGGCGCGAGCGGCGCCAAAGGCGG 8078  
QY 686 TCAAGTTATTCAGCCCGGTAGAGGCAACGGAT-----CGTTTTCTCCAA 730  
Db 8077 GGAACGGGGAACCGGCGCGACGCGCGCGCTGACGCGGCGGATCGCCGTAACGTCTGCA 8018  
QY 731 CAATCTCTGCGGTGACCTTCGCGATGATGAGGCGATGCGCTTCGTCGACCGGCA 790  
Db 8017 CCGCGCGACCTGCGCGCGACGACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7958  
QY 791 --GCGTGAATCGGGGAGATTATGAAACGTTGGGGGACGTCGTAACGTTTCTG 848  
Db 7957 CCGCGCGACGACGCTGACATCGCGCTGCTGCGAGGACTTGTGCAACTGCGCGG 7898  
QY 849 TCGCGGTGTCGCGCATTTAGCCCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCG 908  
Db 7897 CACCGCGATCTCGGGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7838  
QY 909 GCGCGGTGCTGATGAGGCA-----GATCGCGCTTACGCTTGGGCGCGAGAG 960  
Db 7837 ACCCGTGGCGCTGATGAGGCAACGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 7778  
QY 961 CTGCGTTTTCGCGT-----CGTGGCATTTTCGTCGCGCG 995  
Db 7777 CTTCGCGATGCGGATCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7718  
QY 996 CGAGCTAAACCTCGCGATCTGTGCTCATTCGACACTTCGCGCGCGCGCGCGCGCG 1055  
Db 7717 CAACCTCGGCTCGCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 7658  
QY 1056 GCGTGAAGAGCTCGCGATCGCAAGCTCGATTGCGCGCGCGCGCGCGCGCGCGCG 1115  
Db 7657 GCGCGAGAGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7598  
QY 1116 TCCCTTCCCGACCGTTTCGCTATTAATGATGAGGCGCGCGCGCGCGCGCGCG 1175  
Db 7597 GCGCGTGGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7538  
QY 1176 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1235  
Db 7537 GCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7478  
QY 1236 GCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1295  
Db 7477 GCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7418  
QY 1296 AGCGGTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1355  
Db 7417 GAGGCGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7358  
QY 1356 CGAGCATCTTTGTAAGAAATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1406  
Db 7357 CGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7307

## RESULT 9

US-09-107-532A-170  
Sequence 170, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucellette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSER: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Maitland  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arinlelelo, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 893-5007  
TELEFAX: (781) 893-8277  
INFORMATION FOR SEQ ID NO: 170:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 870 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
MOLECULE TYPE: DNA (genomic)  
HYDROTHERMAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...870  
SEQUENCE DESCRIPTION: SEQ ID NO: 170:  
US-09-107-532A-170  
Query Match 7.1%; Score 104.2; DB 4; Length 870;  
Best Local Similarity 50.9%; Pred No. 1.1e-19;  
Matches 247; Conservative 0; Mismatches 238; Indels 0; Gaps 0;  
QY 179 ATGAATCGGTAATTCAGAGAGGCGACCGTGAAGGCGGATCCGCTAATTTTGT 238  
Db 333 ATGAAGCACTCTCACTGTAAGCAACATTCAGGAGTTCCCTCAATTTAGAGTGA 382  
QY 239 CCGATTTTCTCTCTCGCGCGTTTCTTGGGCGCGTCCGCTGCGCATGAAG 298  
Db 383 TGGATCTTAATTTTATCATGGAAGTATGGAACGATTCGAGAAAAAGATCACAGAT 442  
QY 239 CGATTCACCGGCGCGAGAGCGGAACCTCCAGCTGCTGCTCCCTGCTCCGCTGCTG 358  
Db 443 TATTCGAACGAGCAACAAAGAAAGTCTTCAGTGTCTTGTTCACAGCTTCAGGTGCG 502  
QY 359 GCGCATCGAGAGAGACATGAGCTTTTGTATGATGATGATGATGATGATGATGATGAT 418  
Db 503 CACGATGACAGAGAGATCTTTTCTTAATGAGATGAGCAAAATCTCTGCTGCTGCA 562  
QY 419 AGCTCACCGGAGGCGCATTTGCGCTTCTGCTGATTTGGCGCAATCCAGATGAGTG 478  
Db 563 AACGCGATAGCAATGAAAGTCTTTTATCTTAAGTATGTAACGAGCCGACGAGCGAG 622  
QY 479 GCGCATGCGCTCGCGGCTTCATCTGCGCATCTTTCAGTCTTTCGCGAGCCGCGCGCA 538  
Db 623 GATGATCTGCACTTTTCAATGAGAGCGATATTTCTGACAGAACCCCAAGCTTGA 682  
QY 539 TAGTTTCTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 598  
Db 683 TCGGTTTGTGGAAGAGCGCTCATGAGAGAGATCAACAGAGAACTTCAGAGAAATT 742  
QY 599 TCGACAGCGGAGATTTGTTGAAACTGCTGATGATGATGATGATGATGATGATGATGAT 658  
Db 743 TCAAAAGCTGAATCTTATATCTCATGCTTTTGTGATGATGATGATGATGATGATGATGAT 802  
QY 659 AATTG 663

Db 803 AATTG 807

RESULT 10

US-09-252-991A-12365/C  
Sequence 12365, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIORITY FILING DATE: 1999-02-18  
PRIORITY APPLICATION NUMBER: US 60/074,788  
PRIORITY FILING DATE: 1998-02-18  
PRIORITY APPLICATION NUMBER: US 60/094,190  
PRIORITY FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 12365  
LENGTH: 444  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-12365

Query Match 6.6%; Score 97.2; DB 4; Length 444;  
Best Local Similarity 51.6%; Pred. No. 7.3e-18;  
Matches 222; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

251 TCCTCGGCGGCTTCTTTGGGCAAGCGTCCGCTGCGATCATGAAAGCGATTCACCGCG 310  
444 TCATGGGGGGTTCATGAGCGCGCATGCTCGGCAAGCGCTTCTGTCGCAAGCGATGCG 385  
311 CCACAGAGCTGAACTCCCACTGCTGCTCCCTCGCTTCGCGTGTGCGCGCATGACAG 370  
384 CCTCGAAGAGCGCTCCCGCTGATCTCTCCGCTTCGCGCGCGCGCGCGCATGACAG 325  
371 AAGACATGAGCTTTTGTATGATGCTGTCATTAACCGCGCTGTGACGCTCACCGCG 430  
324 AAGCGCTATCTGCTGATGAGATGCGCAAGACCTCGCGCTCTCGCGCGCTCGCG 265  
431 AGGCGCATTTGCGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 490  
264 AAGAGAGCATCCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 205  
491 CGTGGGCTTCACTCTGCGCATCTCACTTTTGCAGAACCGCGCGCGAGATAGCTTCTCG 550  
204 GCTGGCGATGCTCGCGCGAGCGATGCTGCGCGAACCGAGCGCTGATCGGCTTCCGCG 145  
551 GTCCTCGCGTGTGAGTTAACTCACTGCGCATGCGCTTTCAGACGCTGTGCGAGCGCG 610  
144 GTCCCGCGCTGATGAGAGCGAGACGCTCGCGAGAACCTGCGAGAGCTTCCAGCGTAG 85  
611 AGAATTTGGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 670  
84 AGTTCTCTCTTGAAGCATGCGCGCATGATGATGCTGCTGCTGCTGCTGCTGCTGCTG 25  
671 CGGTGCGCAAA 680  
24 GCTGCGCGCA 15

RESULT 11

US-08-961-527-7/C  
Sequence 7, Application US/08961527  
Patent No. 6420135  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESS: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 19702 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-7

Query Match 6.5%; Score 96.2; DB 4; Length 19702;  
Best Local Similarity 49.5%; Pred. No. 7.6e-17;  
Matches 248; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

179 ATGAATGATTAATTAATGAGAGGACCGTGGAGGCGATTCGCGTATTTTGT 238  
11553 ATGAAGCGCTTGTGACAGAGACTGCTTATTAAGGTACAGCTGTGGCTTGGATTA 11494  
239 CGGATTTTCTTCTGCGGCTTCTTTGGGCAAGCGTGGCGTGGCTGCTGCTGCTGCT 298  
11493 TGAATTTTAACTTATATGCTGCTTCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 11434  
299 CGATTCACCGGCGACAGAGGTGAATCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358  
11433 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11374  
359 CGCGCATGAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 418  
11373 CCGCTATGAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 11314  
419 AGCGTACCGGCGAGCGCATTTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 478  
11313 AACCGCATTAAGAGCGCTGCTTCTTAACTGACCATTTGACAGATGACAGATGAGT 11254  
479 GCGCGATGCGCTGCGGCGCTTCACTGCGCATCTCACTTTTGCAGAACCGCGCGCAG 538  
11253 GTGTGACAGCTTCTTGTGCTATGAGAGCGATATCTTGGCTGAAACACAGAGCTTGG 11194  
539 TAGGTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 598  
11193 TTGCTTTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 11134  
599 TCGACAGCGCGAGAAATTTGCTGAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 658  
11133 TCGAAAGCGAGAAATTTCTTATGAAACATGCTTGTGATGCTATTTGTCAAAAAGAG 11074  
659 AATTGCGTGCAGCGGTGCA 679  
11073 ACTTACAGATACATTTGCTA 11053

RESULT 12

US-09-557-884-1  
Sequence 1, Application US/09557884  
Patent No. 6506581

GENERAL INFORMATION:

APPLICANT: Fleischmann et al.  
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments

NUMBER OF SEQUENCES: 1  
Thereof, and Uses Thereof

CORRESPONDENCE ADDRESS:

ADDRESS: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: MD  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS V6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/557,884  
FILING DATE: 25-Apr-2000  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,102  
FILING DATE: JUN-5-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB186P3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-309-8504  
TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-557-884-1

Query Match 6.5%; Score 96.2; DB 4; Length 1830121;  
Best Local Similarity 49.5%; Pred. No. 5.7e-16;  
Matches 248; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

179 ATGATCGGTAATTAATGAGAGGACCGTGGAGGCAATCCGGTAGCCCTTATTTGT 238  
1336980 AAGATGCGCTAATTAATGAGAGGACCGTGGAGGCAATCCGGTAGCCCTTATTTGT 1337039  
239 CGATTTTCTTCTCCGCGGCTTTTGGGACCGGTCGCGTGGTGCATCATGAAG 298  
1337040 GGAATTTTCTTCTCCGCGGCTTTTGGGACCGGTCGCGTGGTGCATCATGAAG 1337099  
299 CGATTCACCGGCGCACAGAGCTGAATCTCCACTGCTGCTCCCTGCTTCCGGTGGTG 358  
1337100 CGGCGAAGAAAGCGATGGAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1337159  
359 CGCGCATGAGGAGAACATGAGCTTTGTCATATGATGCTTCCATACCGGCGGCTGTC 418  
1337160 CTCGATGAGGAGAACATGATTTCTTTAATGCAAAAGCTAGTCCGTAATTC 1337219  
419 AGCGTACCGGAGGCGCATTTGCGCTTCTGCTGATTTTGGGCAATCCGATGGGTG 478  
1337220 CTCGAATGCGTGAAGAGGTGCGCATTTATTCAGTATTAACGATCGACTTAAGCG 1337279  
479 GCGCATGCGCTCGTGGGCTTATCTGGGCAATCTGCTTTGGGAACCGGCGGCGAGA 538  
1337280 GCGTATCAGCAAGTTTTCGATGTTAGGGGATTTAAATATATGCGAGCAAAAGCCTTAA 1337339

539 TAGTTTCTGAGGCTCGCGGTGTGAGTTAACCATGGGCATCGCTTCCAGACGGTG 598  
1337340 TTGTTTTCGAGGCGCACCGGTTATTTGAACAACCTGCGTGAATAATTTCCAGAGTT 1337399  
599 TCGACGAGCGGAGAAATTTGTGAAACTGTGTGATTTGATGAAATTTGTGCGCATCC 658  
1337400 TCCACGTAAGTAGTTTCTACTTGTGAGAAAGGCGCAATGATGATCGTGAACGTTGAG 1337459  
659 AATTCGTCGACCGGTGGCA 679  
1337460 AATTCGTCGCAACTTTAGCAA 1337480

RESULT 13

US-09-643-990A-1  
Sequence 1, Application US/09643990A  
Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann  
Mark D. Adams  
Owen White  
Hamilton O. Smith  
J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments

NUMBER OF SEQUENCES: 1  
Thereof, and Uses Thereof

CORRESPONDENCE ADDRESS:

ADDRESS: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville,  
STATE: MD  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette  
COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS V6.22  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,990A  
FILING DATE: 23-Aug-2000  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429  
FILING DATE: 1995-06-07  
APPLICATION NUMBER: 08/426,787  
FILING DATE: 1995-04-21  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790  
TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1

Query Match 6.5%; Score 96.2; DB 4; Length 1830121;  
Best Local Similarity 49.5%; Pred. No. 5.7e-16;  
Matches 248; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

179 ATGATCGGTAATTAATGAGAGGACCGTGGAGGCAATCCGGTAGCCCTTATTTGT 238  
1336980 AAGATGCGCTAATTAATGAGAGGACCGTGGAGGCAATCCGGTAGCCCTTATTTGT 1337039  
239 CGATTTTCTTCTCCGCGGCTTTTGGGACCGGTCGCGTGGTGCATCATGAAG 298

Db 1337040 CGAATCTTCTTTATGCGCGGTTCATGGCTTCTGTAGTGTGCAAAATTTGTTAAAG 1337099  
Qy 299 CGAATTCACCGCGCCACAGAGCTGAAATCCCACTGCTGTCTCCCTGCTCCGGTGGTG 358  
Db 1337100 CGGCTGTAAGAAAGCGAATGAAATTTGTCATTTGTGTCTCTCGAGTGGTGGTG 1337159  
Qy 359 CGCGCATGAGGAAAGCAATCCAGCTTTGTGATGATGTGTGCCAATACCGGGGCTGTGC 418  
Db 1337160 CTGTGATGAGGAAAGCAATTTCTCTTATATGCAAAATGCAAAACTATGTCGTAATTG 1337219  
Qy 419 AGCGTACCGCGAGCGCATTTGCCGTCCTGTGTATTTGGCGCAATCCACAGATGGTG 478  
Db 1337220 CTCGAATGAGTGAAGGATGTGCCATTTATTCAGTATTAACGATCCGATTTAGGGG 1337279  
Qy 479 GCGGCATGCGCTCTGTGGGTTCATCTGGGCATCTCACTTTTGGGGAACCGCGCGCGACA 538  
Db 1337280 GCGTATCACAGTTTGGGATGTAGGGGATTTAAATATGCGCGCAAAAGCCTTAA 1337339  
Qy 539 TAGGTTTCTGGGTCTCGCGGTGTGAGTTAAACCATGGGATGCGCTTCAGACGGTG 598  
Db 1337340 TTGGTTTTCAGAGCGCCACGCTTATTTGAACAACTGTGCGTGAATAATTCAGAAAGTT 1337399  
Qy 599 TCGACAGCGCGAGATTTGTGAATACTGTGTGATGAGAAATTTGTGCGCACTCC 658  
Db 1337400 TCCACGTAGTGTACTTCTTACTGAGAAAGGGCAATTGATGATCGTGAACCTTCAG 1337459  
Qy 659 AATTGCGTACAGCGGTGCCAA 679  
Db 1337460 AATGCGTCAAACTTACGAA 1337480

## RESULT 14

US-09-347-878-25  
; Sequence 25, Application US/09347878C  
; Patent No. 6376210  
; GENERAL INFORMATION:  
; APPLICANT: Yuan, Chong  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES  
; FILE REFERENCE: 25865-1651  
; CURRENT APPLICATION NUMBER: US/09/347,878C  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 25  
; LENGTH: 2561  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1131)..(2399)  
; OTHER INFORMATION: Escherichia coli nucleic acid encoding  
; OTHER INFORMATION: folypolyglutamate synthetase-dihydrofolate  
; OTHER INFORMATION: synthetase  
; PUBLICATION INFORMATION:  
; DATABASE ACCESSION NUMBER: M32445/Genbank  
US-09-347-878-25

Query Match 6.3%; Score 93.2; DB 4; Length 2561;  
Best Local Similarity 47.9%; Pred. No. 2,1e-16;  
Matches 269; Conservative 0; Mismatches 293; Indels 0; Gaps 0;  
Qy 200 AAGGACCGTGAAGGCGATTCGGGTAGCCGTTATTTTGTCCGATTTTCTTCTCGGCG 259  
Db 481 AAGGACCTCTGTATGAAAGCCGGTGTGCTGCGCATTCAGTTGCTTTATATGAGG 540  
Qy 260 GTTCTTTGGGCAAGGTCGGGTGGGCGCATATGAAGCGAATTCACCGCGCCACAGGC 319  
Db 541 GTTCAATGAGGCTCTGTTGTGGGTGACAGTTTGTGTGTCGCTTACAGAGGCGCTGGAAG 600  
Qy 330 TGAACCTCCCACTGCTGTCTCCCTGTCTTCGGTGTGGCGGCAATGAGAAACAATC 379  
Db 601 ATTAATGCGCGCTGATCTGCTTCTCGGCTCTGTGGGCGCAAGTATGAGAAAGCACTGA 660

Qy 380 GAGCTTTGTATGATGTGTTCATTAACCGGCGCTGTGCAAGCTTCACCGGAGCGCAT 439  
Db 661 TGTGCTGATGCAATGCGGAAACCTGTGGGCACTGGCAAAATGACAGAGCGGCT 720  
Qy 440 TGCCGTTCTGGTGTATTTGGCAATCCACGATGGGTGGCGCATGAGCTCGTGGGGTT 499  
Db 721 TGCCGTATCATCTCGGTGTGACCGACCGAGAGATGGGCGGGTGTTCGCAAGTTTCGCCA 780  
Qy 500 CATGCGCATCTCACTTTTGGGAAACCGGCGCGAGATAGTTTCTGTGGTCTCGCG 559  
Db 781 TGCTGGGCGATCTCAATGCTGAAACGAAAGGTTAATGAGCTTCCGCTCGCGTGT 840  
Qy 560 TGTGTGAGTTAACCACTGGGCGATGCGCTTCCAGACGGTGTGCGACGAGCGGAGATTGG 619  
Db 841 TATGGAAGAGCCGTTCGCGAAACATGCGCGCTGTGATTCAGCGAGATGAATTCCTGA 900  
Qy 620 TGAACCTGTGTATGATGATGATGATGTCGCACTCAATTGCGTGCAGCGGTGCCAA 679  
Db 901 TCGAAGAAAGGCGCATGCAATGATCGTGTGCTCGGAAATGCGCTGAAACTGGCGA 960  
Qy 680 AAACCTCAAGTTATTCAGCGGTGAGGCAAGCATCGTTTTCGCAACAACTCCTG 739  
Db 961 GCATTCGCGCAAGTGTGATCTGCCAGGCGCAATCTGAAGCGCGCGGTGAAGGCG 1020  
Qy 740 GCGTGGCACTTCGGGTATGGA 761  
Db 1021 TAGTGTACCCCGGTACCGGA 1042

## RESULT 15

US-08-973-275-2  
; Sequence 2, Application US/08973275B  
; Patent No. 5958706  
; GENERAL INFORMATION:  
; APPLICANT: MATSUNAGA, Tadashi  
; APPLICANT: KAMIYA, Shinji  
; APPLICANT: NAKA, Kenryo  
; TITLE OF INVENTION: FINE MAGNETIC PARTICLES CONTAINING USEFUL PROTEINS  
; TITLE OF INVENTION: BOUND THERETO, PROCESS FOR PRODUCING THE SAME, AND USE  
; FILE REFERENCE: MATSUNAGA  
; CURRENT APPLICATION NUMBER: US/08/973,275B  
; EARLIER FILING DATE: 1998-02-09  
; EARLIER APPLICATION NUMBER: PCT/JP97/01043  
; EARLIER FILING DATE: 1997-03-27  
; EARLIER APPLICATION NUMBER: JP 8-97536  
; EARLIER FILING DATE: 1996-03-28  
; EARLIER APPLICATION NUMBER: JP 8-146833  
; EARLIER FILING DATE: 1996-05-16  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 954  
; TYPE: DNA  
; ORGANISM: Magnetospirillum AMB-1  
; FEATURE:  
; NAME/KEY: gene  
; LOCATION: (217)..(702)  
; OTHER INFORMATION: Predicted region encoding membrane bound portion  
; OTHER INFORMATION: of the mps gene product. "n" is any of a, t, c or  
; OTHER INFORMATION: g.  
US-08-973-275-2

Query Match 5.9%; Score 87.4; DB 2; Length 954;  
Best Local Similarity 33.6%; Pred. No. 5.8e-15;  
Matches 152; Conservative 54; Mismatches 247; Indels 0; Gaps 0;  
Qy 951 GCGGAGAGCTGCTTTTGGCGGTGCGCATTTTGGCGCGCGAGCTAACTGCGC 1010  
Db 393 RCGAGAGGTATYAGNABRCNABRMGNYNATGAGATGCGAGAYCAATYYCARGTNC 452  
Qy 1011 GATGTGTCAATATGACACCTTCGCGCGCGAATGTGTGCAAGCGGCTAGAGAGCTCGG 1070



Db 453 NATHTACNNTGTGTGAACGCGNGGNCNTAYCCNGGNGTGTGA YGCNGARGCNMNGG 512  
QY 1071 CATGCCAAGCTGATTTGGCGGCACCTTGTCCAAAGCTTATCGACGCTCCCTCCCAACCGT 1130  
Db 513 NCARGCNGARGCNATHGCGMGNMNSNATHGARA CNTGYTMAAYGTMMGNGTNCNVTNGT 572  
QY 1131 TTCGCTCATTTATTTGTCAGGGCGCTTGGCGGCGCTGCGCATGCTGCCGCCGATCT 1190  
Db 573 NMSNGTNTNATHATGCGNGARGNGGNGMNSGNGGNCNATHGCVTTTNGCNACNGNAAYAC 632  
QY 1191 GGTCTACGCGGCGGAAAACGCGTGGCTGTCCGCATTGCCACGAGAGGCGCCTCGGCCAT 1250  
Db 633 NGTNTNATGTGTNGARCA YGCNATHATAYMSNGTNTATHMSNCCNGARGGNTGYGCNWSNAT 692  
QY 1251 CCTCTCGCGGACACCAACGCGCGCGGAAATCATAGCGGACGCAAGCGCTGCAGCGCA 1310  
Db 693 HYTNTGGMGNMNSNGCNAARAYGCNAGCNGARCAR YTMGNTNACNCGNCA 752  
QY 1311 CGCACTTTTAAAGCCAAAGGCTTATCGACGGGATCGTCCGGAACCGAGCACTTTGTTGA 1370  
Db 753 RGAYYTNCAVAAAYTNMNSNATHATGATWSNGTNGTNCNGARCCNATGGGNGGNCNCA 812  
QY 1371 AGAAATTTCTCGGCACAATCGACAACGCGCTCTC 1403  
Db 813 YMGNAAYCCNGAYYTNATGATGCARACNVTNMS 845

Search completed: November 13, 2003, 00:15:17  
Job time : 124.346 secs

**THIS PAGE BLANK (USPTO)**



Db 1 GTGAGAAAGCGTTTCCGACTATGATGATGAGGAGCAACAATTGACGCTC 60  
QY ATAGACTGCGTTTGGAGCCCTGACAGCTTCACTTCTTGGAAATGAACCTCCCAATATGAC 120  
Db 61 ATAGACTGCGTTTGGAGCCCTGACAGCTTCACTTCTTGGAAATGAACCTCCCAATATGAC 120  
QY 121 AACCTCAATCAAGGCTATGACAGAGCCTTGGAGCGGCTCGAAGAGAGCCCAATATGAC 180  
Db 121 AACCTCAATCAAGGCTATGACAGAGCCTTGGAGCGGCTCGAAGAGAGCCCAATATGAC 180  
QY 181 GAATGGGTAAATTAATGAGAGAGAGCCTTGGAGCGGCTCGAAGAGAGCCCAATATGAC 240  
Db 181 GAATGGGTAAATTAATGAGAGAGAGCCTTGGAGCGGCTCGAAGAGAGCCCAATATGAC 240  
QY 241 GATTTTCTCTCTGCGGCGGTTCTTGGAGAGAGCTGCGTGGTGGAGATCAAGAGAGG 300  
Db 241 GATTTTCTCTCTGCGGCGGTTCTTGGAGAGAGCTGCGTGGTGGAGATCAAGAGAGG 300  
QY 301 ATTCAACCGCCACAGAGCTGAATCTCCCACTGCTGCTCCCTGCTTCCGCTGCTGCG 360  
Db 301 ATTCAACCGCCACAGAGCTGAATCTCCCACTGCTGCTCCCTGCTTCCGCTGCTGCG 360  
QY 361 CGCATGAGAGAGAGCAATGAGCTTGTGATGATGATGATGATGATGATGATGATGATG 420  
Db 361 CGCATGAGAGAGAGCAATGAGCTTGTGATGATGATGATGATGATGATGATGATGATG 420  
QY 421 CGTCAACCGAGAGAGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
Db 421 CGTCAACCGAGAGAGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
QY 481 GGCATGAGAGAGAGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 540  
Db 481 GGCATGAGAGAGAGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 540  
QY 541 GGTTCCTGAGAGAGAGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
Db 541 GGTTCCTGAGAGAGAGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
QY 601 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
Db 601 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
QY 661 TTGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
Db 661 TTGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
QY 721 TTTTCTCCAACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
Db 721 TTTTCTCCAACTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
QY 781 GACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
Db 781 GACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
QY 841 CTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
Db 841 CTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
QY 901 GGGGGGCGGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
Db 901 GGGGGGCGGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
QY 961 CTGGCTTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Db 961 CTGGCTTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
QY 1021 ATCATGACACCTCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
Db 1021 ATCATGACACCTCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
QY 1081 TCGATTTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
Db 1081 TCGATTTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140

QY 1141 ATTGTCAGAGAGCGTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
Db 1141 ATTGTCAGAGAGCGTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
QY 1201 GCCGAAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
Db 1201 GCCGAAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260  
QY 1261 GACACCAACAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
Db 1261 GACACCAACAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320  
QY 1321 AGCCAAAGGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
Db 1321 AGCCAAAGGCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
QY 1381 GGCACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
Db 1381 GGCACATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440  
QY 1441 AGTGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1473  
Db 1441 AGTGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1473

RESULT 2  
US-10-024-370-2  
; Sequence 2, Application US/10024370  
; Publication No. US20020142405A1  
; GENERAL INFORMATION:  
; APPLICANT: TIIG, YVONNE  
; APPLICANT: ELMANN, BERND  
; APPLICANT: EGGELING, LOTAR  
; APPLICANT: SAHM, HERMANN  
; APPLICANT: MOCKEL, BETTINA  
; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY  
; TITLE OF INVENTION: FERMENTATION AND NUCLEOTIDE SEQUENCES CODING FOR THE  
; FILE REFERENCE: 21123-284139-MAS  
; CURRENT FILING DATE: US/10/024,370  
; PRIOR APPLICATION NUMBER: 2001-12-21  
; PRIOR FILING DATE: 1999-07-29  
; PRIOR APPLICATION NUMBER: DE 199 24 365.4  
; PRIOR FILING DATE: 1999-07-29  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 1473  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1) ..(1473)  
; OTHER INFORMATION: accda  
US-10-024-370-2

Query Match 100.0%; Score 1473; DB 13; Length 1473;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAGAAAGCGTTTCCGACTATGATGATGAGGAGCAACAATTGACGCTC 60  
Db 1 GTGAGAAAGCGTTTCCGACTATGATGATGAGGAGCAACAATTGACGCTC 60  
QY 61 ATAGACTGCGTTTGGAGCCCTGACAGCTTCACTTCTTGGAAATGAACCTCCCAATATGAC 120  
Db 61 ATAGACTGCGTTTGGAGCCCTGACAGCTTCACTTCTTGGAAATGAACCTCCCAATATGAC 120  
QY 121 AACCTCAATCAAGGCTATGACAGAGCCTTGGAGCGGCTCGAAGAGAGCCCAATATGAC 180  
Db 121 AACCTCAATCAAGGCTATGACAGAGCCTTGGAGCGGCTCGAAGAGAGCCCAATATGAC 180

QY 181 GAATCGTAATTAATGAGAGAGCAGCGTGAAGGATTCGGTACCGCTTAATTTGTCC 240  
 DB 181 GATGCGTAATTAATGAGAGAGCAGCGTGAAGGATTCGGTACCGCTTAATTTGTCC 240  
 QY 241 GATTTTCTTCTCTCGCGGCTTTTGGGACCGTCCGTCGGTGCATCATGAAGGCG 300  
 DB 241 GATTTTCTTCTCTCGCGGCTTTTGGGACCGTCCGTCGGTGCATCATGAAGGCG 300  
 QY 301 ATTCACCGGCGCAGAGAGTGAAGCTCCAGTGCCTGCTCCCGCTTCGGTGTGGC 360  
 DB 301 ATTCACCGGCGCAGAGAGTGAAGCTCCAGTGCCTGCTCCCGCTTCGGTGTGGC 360  
 QY 361 CCATGACAGAGAGACATGAGCTTTTGCATGATGATGATCAACCGCGCTGTCAG 420  
 DB 361 CCATGACAGAGAGACATGAGCTTTTGCATGATGATGATCAACCGCGCTGTCAG 420  
 QY 421 CGTACCGGCGCAGAGAGTGAAGCTCCAGTGCCTGCTCCCGCTTCGGTGTGGC 480  
 DB 421 CGTACCGGCGCAGAGAGTGAAGCTCCAGTGCCTGCTCCCGCTTCGGTGTGGC 480  
 QY 481 GCGATGCGCTCGTGGGCTCATCTGGGCACTCACTTTTGGGAAACCGCGCGCAGAT 540  
 DB 481 GCGATGCGCTCGTGGGCTCATCTGGGCACTCACTTTTGGGAAACCGCGCGCAGAT 540  
 QY 541 GGTTCCTGCGTCTGCGCTGAGTAACTCACTGGGCACTGCGCTTCAGACGGTGTG 600  
 DB 541 GGTTCCTGCGTCTGCGCTGAGTAACTCACTGGGCACTGCGCTTCAGACGGTGTG 600  
 QY 601 CAGCAGCGGAGAAATTTGTGTAAGAACTGTGTGATGATGATGATGATGATGATGAT 660  
 DB 601 CAGCAGCGGAGAAATTTGTGTAAGAACTGTGTGATGATGATGATGATGATGATGAT 660  
 QY 661 TTGCGTGAAGCGGCTGCGCAAAACCTCAAGTATTAACCGGTGAGGCAACGATCGT 720  
 DB 661 TTGCGTGAAGCGGCTGCGCAAAACCTCAAGTATTAACCGGTGAGGCAACGATCGT 720  
 QY 721 TTTTCTCCAACTCTGCGCTGAGCACTTCGCGTGAAGGAGGCACTTCGCGTCTCGT 780  
 DB 721 TTTTCTCCAACTCTGCGCTGAGCACTTCGCGTGAAGGAGGCACTTCGCGTCTCGT 780  
 QY 781 GACCGCGAGAGGCTGGAATCGGAGAGATTAAGAACTGTGGGCGAGACGTCGTCAAG 840  
 DB 781 GACCGCGAGAGGCTGGAATCGGAGAGATTAAGAACTGTGGGCGAGACGTCGTCAAG 840  
 QY 841 CTTTCTGCGCGCTGCGCTGAGCACTTCGCGTGAAGGAGGCACTTCGCGTCTCGT 900  
 DB 841 CTTTCTGCGCGCTGCGCTGAGCACTTCGCGTGAAGGAGGCACTTCGCGTCTCGT 900  
 QY 901 GGGGCGCGGCGCTGAGTGAAGGAGAGTTCGCGCTTCAGCGCTTGGGCGGAGAG 960  
 DB 901 GGGGCGCGGCGCTGAGTGAAGGAGAGTTCGCGCTTCAGCGCTTGGGCGGAGAG 960  
 QY 961 CTGCGCTTTCGCGCTGAGTGAAGGAGAGTTCGCGCTTCAGCGCTTGGGCGGAGAG 1020  
 DB 961 CTGCGCTTTCGCGCTGAGTGAAGGAGAGTTCGCGCTTCAGCGCTTGGGCGGAGAG 1020  
 QY 1021 ATGATCGACACTCCGCGCGCGCAATTTGTGAGGCGGCTGAGAGGCTCGGCAATCGCAAGC 1080  
 DB 1021 ATGATCGACACTCCGCGCGCGCAATTTGTGAGGCGGCTGAGAGGCTCGGCAATCGCAAGC 1080  
 QY 1081 TCGATTTGCGCGCACTTTCGCAAGCTTTCAGCGCTTCGCGCTTCGCGCTTCGCGCTTC 1140  
 DB 1081 TCGATTTGCGCGCACTTTCGCAAGCTTTCAGCGCTTCGCGCTTCGCGCTTCGCGCTTC 1140  
 QY 1141 ATTGATCAAGGCGGCTGAGGAGGAGTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTC 1200  
 DB 1141 ATTGATCAAGGCGGCTGAGGAGGAGTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTC 1200  
 QY 1201 GCGGAAACGCGGCTGAGGAGGAGTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTC 1260  
 DB 1201 GCGGAAACGCGGCTGAGGAGGAGTTCGCGCTTCGCGCTTCGCGCTTCGCGCTTC 1260

QY 1261 GACACCAACCAACCGCGGAAATCATAGAGCAGACGAGCGTGCAGGCGCACACTTTTA 1320  
 DB 1261 GACACCAACCAACCGCGGAAATCATAGAGCAGACGAGCGTGCAGGCGCACACTTTTA 1320  
 QY 1321 AGCCAAAGGCTTATGACGCGGATGCTCCGCGAAACCGAGCACTTTGTAAGAAATTC 1380  
 DB 1321 AGCCAAAGGCTTATGACGCGGATGCTCCGCGAAACCGAGCACTTTGTAAGAAATTC 1380  
 QY 1381 GGCACAAATCAGAACGCGCTTCGCGAAATGATCAATCCGAGAGAGGCGGAGCGGAC 1440  
 DB 1381 GGCACAAATCAGAACGCGCTTCGCGAAATGATCAATCCGAGAGAGGCGGAGCGGAC 1440  
 QY 1441 AGTCGCTTCACAGATTTGAGCGTTTACGCGAG 1473  
 DB 1441 AGTCGCTTCACAGATTTGAGCGTTTACGCGAG 1473  
 RESULT 3  
 US-10-024-370-1  
 ; Sequence 1, Application US/10024370  
 ; Publication No. US20020142405A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TILG, YVONNE  
 ; APPLICANT: ELKMANNS, BERND  
 ; APPLICANT: EGSELING, LOTHAR  
 ; APPLICANT: SAHM, HERMANN  
 ; APPLICANT: MCKEL, BETTINA  
 ; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY  
 ; TITLE OF INVENTION: FERMENTATION AND NUCLEOTIDE SEQUENCES CODING FOR THE  
 ; FILE REFERENCE: 21123-284139-MAS  
 ; CURRENT APPLICATION NUMBER: US/10/024,370  
 ; PRIOR FILING DATE: 2001-12-21  
 ; PRIOR APPLICATION NUMBER: 09/362,899  
 ; PRIOR FILING DATE: 1999-07-29  
 ; PRIOR APPLICATION NUMBER: DE 199 24 365.4  
 ; NUMBER OF SEQ ID NOS: 3  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 2123  
 ; TYPE: DNA  
 ; ORGANISM: Corynebacterium glutamicum  
 ; FEATURE:  
 ; NAME/KEY: gene  
 ; LOCATION: (508)..(1980)  
 ; OTHER INFORMATION: accda  
 US-10-024-370-1  
 Query Match 100.0%; Score 1473; DB 13; Length 2123;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTGAGAGCGTTTCCGATGATGATGAGGAGCAATGAAACACTTCAGATTAATGAGCTC 60  
 DB 508 GTGAGAGCGTTTCCGATGATGATGAGGAGCAATGAAACACTTCAGATTAATGAGCTC 60  
 QY 61 ATGAGCTCGTTTGGAGCCTGACAGCTTCATTTCTTGAATGAATCTCCCAATATGAC 120  
 DB 568 ATGAGCTCGTTTGGAGCCTGACAGCTTCATTTCTTGAATGAATCTCCCAATATGAC 120  
 QY 121 AACCTCAATCAAGGCTATGAGAGACCTTGGAGGCGGCTGCAACCAAGGCCAATGCGAT 180  
 DB 568 AACCTCAATCAAGGCTATGAGAGACCTTGGAGGCGGCTGCAACCAAGGCCAATGCGAT 180  
 QY 628 AACCTCAATCAAGGCTATGAGAGACCTTGGAGGCGGCTGCAACCAAGGCCAATGCGAT 687  
 DB 181 GAATCGTAATTAATGAGAGAGCAGCGTGAAGGAGCAATTCGATAGCGCTTAATTTGTCC 240  
 DB 688 GAATCGTAATTAATGAGAGAGCAGCGTGAAGGAGCAATTCGATAGCGCTTAATTTGTCC 240  
 QY 241 GATTTTCTTCTCTCGCGGCTTTCTTTGGGACCGGTCGCGTGCATCATGAAGGCG 300  
 DB 748 GATTTTCTTCTCTCGCGGCTTTCTTTGGGACCGGTCGCGTGCATCATGAAGGCG 807

Qy 301 ATTCACCGCGCCAGAGAGTGAATCTCCAGTGTGTCTCCCTGCTTCCGATGTGTCG 360  
Db 808 ATTACCGCGCCAGAGAGTGAATCTCCAGTGTGTCTCCCTGCTTCCGATGTGTCG 867  
Qy 361 CGCATGCAAGAAAGCAATGAGCTTTTGTCAATGATGTGTCCATPAACCGCGCTGTGAG 420  
Db 868 CGCATGCAAGAAAGCAATGAGCTTTTGTCAATGATGTGTCCATPAACCGCGCTGTGAG 927  
Qy 421 CGTACCGCGAGGCGCATTTTCCGTTCCGTGTGATTTTGGCCATCCCAAGATGGGTGGC 480  
Db 928 CGTACCGCGAGGCGCATTTTCCGTTCCGTGTGATTTTGGCCATCCCAAGATGGGTGGC 987  
Qy 481 GGCATGCGCTGTGAGGCTTCACTGTGGCATCTCACTTTTGGGAAACCGCGCGCAGATA 540  
Db 988 GGCATGCGCTGTGAGGCTTCACTGTGGCATCTCACTTTTGGGAAACCGCGCGCAGATA 1047  
Qy 541 GGTTCCTGTGGTCTCTCGCGTGTGAGTTAAACCATGCGGCATGCGCTTCCAGCGGTGTG 600  
Db 1048 GGTTCCTGTGGTCTCTCGCGTGTGAGTTAAACCATGCGGCATGCGCTTCCAGCGGTGTG 1107  
Qy 601 CAGGAGCGGAGAAATTTGTGTAACCTGTGTGATTTGGAATTTGTGTGCGCACTCCAA 660  
Db 1108 CAGGAGCGGAGAAATTTGTGTAACCTGTGTGATTTGGAATTTGTGTGCGCACTCCAA 1167  
Qy 661 TTGGGTGAGCGGTGTGCAAAAACCTCAAGTTATTCAAGCGGTGAGAGCAACGATGCT 720  
Db 1168 TTGGGTGAGCGGTGTGCAAAAACCTCAAGTTATTCAAGCGGTGAGAGCAACGATGCT 1227  
Qy 721 TTTTCTCCAACTCTCTGCGGTGACCTTCGCGTGTATGAGAGCGATTTGCGCTTCTGT 780  
Db 1228 TTTTCTCCAACTCTCTGCGGTGACCTTCGCGTGTATGAGAGCGATTTGCGCTTCTGT 1287  
Qy 781 GACCGGAGAGCGCTGGAATGCGGGAGATTATGAAAGTTGGGGGCGAGAGTGTGTCAAG 840  
Db 1288 GACCGGAGAGCGCTGGAATGCGGGAGATTATGAAAGTTGGGGGCGAGAGTGTGTCAAG 1347  
Qy 841 CTTTCTGTGTGCGGTGTGCGCAATTGAGCCCGCTGTGCGCGTTCCTTGGCGCGCATC 900  
Db 1348 CTTTCTGTGTGCGGTGTGCGCAATTGAGCCCGCTGTGCGCGTTCCTTGGCGCGCATC 1407  
Qy 901 GGGGGCGCGCGGTGTGTCTGATTTGGGAGATGCGCGCTTCAAGCTTGTGGCGCGAGAG 960  
Db 1408 GGGGGCGCGCGGTGTGTCTGATTTGGGAGATGCGCGCTTCAAGCTTGTGGCGCGAGAG 1467  
Qy 961 CTGGTTTGTGCGGTGTGTGCGCAATTGCGTGGCGCGAGCTTAACCTGCGATGTGTGCC 1020  
Db 1468 CTGGTTTGTGCGGTGTGTGCGCAATTGCGTGGCGCGAGCTTAACCTGCGATGTGTGCC 1527  
Qy 1021 ATCATGCAACCTCTCGCGCGCGCAATTGTGCGAGCGGCTGAGAGCTGCGCATGCGCAAGC 1080  
Db 1528 ATCATGCAACCTCTCGCGCGCGCAATTGTGCGAGCGGCTGAGAGCTGCGCATGCGCAAGC 1587  
Qy 1081 TCGATTTGGCGCACTTGTCCAGTTATTCAGCTTCCCTTCCCACTGTTCCGTCAATT 1140  
Db 1588 TCGATTTGGCGCACTTGTCCAGTTATTCAGCTTCCCTTCCCACTGTTCCGTCAATT 1647  
Qy 1141 ATTGGTCAAGGCGTTGGCGGTGCGCGCTGGCAATGCTGCGCGCGATCTGTCTAAGCGG 1200  
Db 1648 ATTGGTCAAGGCGTTGGCGGTGCGCGCTGGCAATGCTGCGCGCGATCTGTCTAAGCGG 1707  
Qy 1201 GCGGAAAACGCGTGTGTGCGCAATTGCGCACAGAGGCGCTGCGCATCTTCCGC 1260  
Db 1708 GCGGAAAACGCGTGTGTGCGCAATTGCGCACAGAGGCGCTGCGCATCTTCCGC 1767  
Qy 1261 GACACCAACACGCGCGGAAATCATAGAGGACAGGCGGTGACGCGCACTTTTA 1320  
Db 1768 GACACCAACACGCGCGGAAATCATAGAGGACAGGCGGTGACGCGCACTTTTA 1827  
Qy 1321 AGCGAAGGCGTTATGAGAGGATGTCGCCGAAACCGAGCACTTTTGAAGAAATTTTC 1380  
Db 1828 AGCGAAGGCGTTATGAGAGGATGTCGCCGAAACCGAGCACTTTTGAAGAAATTTTC 1887  
Qy 1381 GGCACAATCAGCAACGCGCTCTCCGAATTGATTAACATCCGAGAGGCGGAGCGGAC 1440

Db 1888 GGCACAATCAGCAACGCGCTCTCCGAATTGATTAACATCCGAGAGGCGGAGCGGAC 1947  
Qy 1441 AGTGGCTTCAAGATTTGAGCGTTTACGCCAG 1473  
Db 1948 AGTGGCTTCAAGATTTGAGCGTTTACGCCAG 1980

RESULT 4  
US-09-738-626-1/c  
; Sequence 1, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patentin ver. 3.0  
; SEQ ID NO 1  
; LENGTH: 3309400  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

Query Match 100.0%; Score 1473; DB 10; Length 3309400;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGAGAGAGCGTTTCCGACTATGTGTGGGCGATGGAACAACCTTACGATTGACGCTC 60  
Db 881114 GTGAGAGAGCGTTTCCGACTATGTGTGGGCGATGGAACAACCTTACGATTGACGCTC 881055  
Qy 61 ATGACTCGGTTTGAACCTGACAGCTTCACTTTTGAATGAATCTCCCAATATGAC 120  
Db 881054 ATGACTCGGTTTGAACCTGACAGCTTCACTTTTGAATGAATCTCCCAATATGAC 880995  
Qy 121 AACCTCAATCAGGCTATGAGAGACCTTGAGCGGCTGGAAGCAAGGCAATGCGAT 180  
Db 880994 AACCTCAATCAGGCTATGAGAGACCTTGAGCGGCTGGAAGCAAGGCAATGCGAT 880935  
Qy 181 GAATCGGTATTAATGAGAAAGCACCTGTGAGAGGCAATTCGGTACCGCTTATTTTGTCC 240  
Db 880934 GAATCGGTATTAATGAGAAAGCACCTGTGAGAGGCAATTCGGTACCGCTTATTTTGTCC 880875  
Qy 241 GATTTTCTTCTCGCGCGGTTCTTTGGGAGCGGTGCGCGGAGGCGCATGAGAGGCG 300  
Db 880874 GATTTTCTTCTCGCGCGGTTCTTTGGGAGCGGTGCGCGGAGGCGCATGAGAGGCG 880815  
Qy 301 ATTCACCGCGCCAGAGCTGAATCTCCACTGTGTCTCCCTGCTTCCGATGTGTCG 360  
Db 880814 ATTCACCGCGCCAGAGCTGAATCTCCACTGTGTCTCCCTGCTTCCGATGTGTCG 880755  
Qy 361 CGCATGCAAGAAAGCAATGAGCTTTTGTCAATGATGTGTCCATPAACCGCGCTGTGAG 420  
Db 880754 CGCATGCAAGAAAGCAATGAGCTTTTGTCAATGATGTGTCCATPAACCGCGCTGTGAG 880695

QY 421 GGTCAACCGGAGCGCAATTGCGCTTCTGATGATTTGCGCAATCCAGATGGGTGC 480  
 DB 880694 CGTCAACCGGAGCGCAATTGCGCTTCTGATGATTTGCGCAATCCAGATGGGTGC 880635  
 QY 481 GCGATGGCTCTGCGGGGTTCACTGCGCATCTCACTTTGCGGAACCGCGCGCGCATG 540  
 DB 880634 GCGATGGCTCTGCGGGGTTCACTGCGCATCTCACTTTGCGGAACCGCGCGCGCATG 880575  
 QY 541 GGTTCCTGAGGCTCGCGTGGAGTGAACAACCTGGGCGATGCGCTTCAGACGGTGG 600  
 DB 880574 GGTTCCTGAGGCTCGCGTGGAGTGAACAACCTGGGCGATGCGCTTCAGACGGTGG 880515  
 QY 601 CAGCAGCGGAGAAATTGTGTAACCTGTGTGATGATGATGATGATGATGATGATGATG 660  
 DB 880514 CAGCAGCGGAGAAATTGTGTAACCTGTGTGATGATGATGATGATGATGATGATGATG 880455  
 QY 661 TTGCGTGAAGCGGTGGCAAAAACCTCAAGGTTATTGACCGGTGAGGCAACGATCGT 720  
 DB 880454 TTGCGTGAAGCGGTGGCAAAAACCTCAAGGTTATTGACCGGTGAGGCAACGATCGT 880395  
 QY 721 TTTTCTCAACCACTCTGCGGTGGCACTTGGGTGATGAGGCGATTTGGCGGTCTGCT 780  
 DB 880394 TTTTCTCAACCACTCTGCGGTGGCACTTGGGTGATGAGGCGATTTGGCGGTCTGCT 880335  
 QY 781 GACCCGAGAGGCTGGAATCGGGAGATTAATGAAACGTTGGGGCAGACGTCGTCAAG 840  
 DB 880334 GACCCGAGAGGCTGGAATCGGGAGATTAATGAAACGTTGGGGCAGACGTCGTCAAG 880275  
 QY 841 CTTTCTGCTGCGGTGCTGCGCATTTGAGCCCGGCTGTGCGGCTTGGCGCGCATC 900  
 DB 880274 CTTTCTGCTGCGGTGCTGCGCATTTGAGCCCGGCTGTGCGGCTTGGCGCGCATC 880215  
 QY 901 GGGGGCGCGCGCTGTGCTGATTTGGGAGATGCCCGCTTACCGCTTGGCGCGAGAG 960  
 DB 880214 GGGGGCGCGCGCTGTGCTGATTTGGGAGATGCCCGCTTACCGCTTGGCGCGAGAG 880155  
 QY 961 CTGCGTTTGGCGCTGTGCGCATTTGCGTGGCGCGCGAGCTTAACCTGCGCATCTGTC 1020  
 DB 880154 CTGCGTTTGGCGCTGTGCGCATTTGCGTGGCGCGCGAGCTTAACCTGCGCATCTGTC 880095  
 QY 1021 ATCATGCACTCTCGGCGCGCAATTGTGCGAGCGCGCTGAGAGCTCGGCGATGCGAAC 1080  
 DB 880094 ATCATGCACTCTCGGCGCGCAATTGTGCGAGCGCGCTGAGAGCTCGGCGATGCGAAC 880035  
 QY 1081 TCGATTTGGCGCATCTTGTCAAGCTTATCGAGCTTCCCTCCCAACCGTTTGGGTCAAT 1140  
 DB 880034 TCGATTTGGCGCATCTTGTCAAGCTTATCGAGCTTCCCTCCCAACCGTTTGGGTCAAT 879975  
 QY 1141 ATTGCTCAGGGCGTTGGCGGTGGCGCGCTGCGCATGCTGCGCGCATCTGCTACGCG 1200  
 DB 879974 ATTGCTCAGGGCGTTGGCGGTGGCGCGCTGCGCATGCTGCGCGCATCTGCTACGCG 879915  
 QY 1201 GCGGAAAACGCGTGTCTGCGCATTTGCCACGAGGGGCGCTGCGCATCTCTTCCGC 1260  
 DB 879914 GCGGAAAACGCGTGTCTGCGCATTTGCCACGAGGGGCGCTGCGCATCTCTTCCGC 879855  
 QY 1261 GACACCAACCAACCGCGGGAATCATAGAGCGAAGGCGTGCAGCGCCACGCACTTTTA 1320  
 DB 879854 GACACCAACCAACCGCGGGAATCATAGAGCGAAGGCGTGCAGCGCCACGCACTTTTA 879795  
 QY 1321 AGCCAAAGGCTTATGAGCGGATCTGCGCGCAACCGAGCACTTGTGGAAGAAATTTTC 1380  
 DB 879794 AGCCAAAGGCTTATGAGCGGATCTGCGCGCAACCGAGCACTTGTGGAAGAAATTTTC 879735  
 QY 1381 GGCACATCAGCAACGCTTCTCGAATTGATTAACAATCCGAGAGGGCGGAGCGGAC 1440  
 DB 879734 GGCACATCAGCAACGCTTCTCGAATTGATTAACAATCCGAGAGGGCGGAGCGGAC 379675  
 QY 1441 AGTCGCTTCAACGATTTGAGGCTTAAAGCGAG 1473  
 DB 879674 AGTCGCTTCAACGATTTGAGGCTTAAAGCGAG 879642

RESULT 5  
 US-09-712-363-35  
 ; Sequence 35, Application US/09712363  
 ; Patent No. US20020164588A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Eiseberg, David  
 ; APPLICANT: Rotstein, Sergio H.  
 ; APPLICANT: Marcotte, Edward M.  
 ; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
 ; FILE REFERENCE: 07419-032001  
 ; CURRENT APPLICATION NUMBER: US/09/712,363  
 ; PRIOR FILING DATE: 2000-11-13  
 ; PRIOR APPLICATION NUMBER: PCT/US00/02246  
 ; PRIOR FILING DATE: 2000-01-28  
 ; PRIOR APPLICATION NUMBER: 60/179,531  
 ; PRIOR FILING DATE: 2000-02-01  
 ; PRIOR APPLICATION NUMBER: 60/117,844  
 ; PRIOR FILING DATE: 1999-01-29  
 ; PRIOR APPLICATION NUMBER: 60/118,206,  
 ; PRIOR FILING DATE: 1999-02-01  
 ; PRIOR APPLICATION NUMBER: 60/126,593  
 ; PRIOR FILING DATE: 1999-03-26  
 ; PRIOR APPLICATION NUMBER: 60/134,093  
 ; PRIOR FILING DATE: 1999-05-14  
 ; PRIOR APPLICATION NUMBER: 60/134,092  
 ; PRIOR FILING DATE: 1999-05-14  
 ; PRIOR APPLICATION NUMBER: 60/165,124  
 ; PRIOR FILING DATE: 1999-11-12  
 ; PRIOR APPLICATION NUMBER: 60/165,086  
 ; PRIOR FILING DATE: 1999-11-12  
 ; NUMBER OF SEQ ID NOS: 292  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 35  
 ; LENGTH: 1488  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 US-09-712-363-35  
 Query Match 21.2%; Score 312.6; DB 10; Length 1488;  
 Best Local Similarity 54.0%; Pred. No. 5.5e-87;  
 Matches 717; Conservative 0; Mismatches 589; Indels 21; Gaps 3;  
 QY 65 ACTGGTTTGGACCTTGACAGCTTTCATTTCTGGAATGAACCTCCCAATATGCAACC 124  
 DB 32 AGCGGTCTAGACCGGGGATCTTTCGACCTGGAGATGAGAGCGCTGGCGGTGCGG 91  
 QY 125 TCAATCAAGGCTATGACAGACTTGTGAGCGGGCTCGAAGCAAGCCAAATGCGATGAT 184  
 DB 92 TAGCGACTCTATGCGGGAGAGCTGGCGCGCTCGGGCGGCAACCGGCGGAGCAT 151  
 QY 185 CGGTAAATTAATGAGAAAGCAGCGTGAAGGCAATTCGGTAGCCGTTATTTTTCGATT 244  
 DB 152 CGGTGACAGACCGGTGAGGAGCGCTATTCGGCGCGGCGGCTGCTGTGAGT 211  
 QY 245 TTTCTCTGCGGGGTTCTTTGGGCAAGGTGCGGTGCGTGCATCATGAAGCGATTC 304  
 DB 212 TCGACTTCTGCGGCTCGATGAGGTGCGAGCGCGCAAGATCAACGCGCGCTGCG 271  
 QY 305 ACCGCGCAGACAGACTGAACTCCCACTGCTGATCTCCCTGCTTCCGCTGCGGCGCA 364  
 DB 272 AGCGGCGACCGCGAGCGGCTGCGCTCATCTGCGGTCAACAGCTCGGAGGAGCGCA 331  
 QY 365 TCGAGAAAGACATGAGCTTTGTATGATGATGATGATGATGATGATGATGATGATG 424  
 DB 332 TCGAAGAGGACAGCGTGGCTTTTGTGAGATGAGGAGATGCTGCGGCTCATCAGCTGC 391  
 QY 425 ACCGAGGCGCATTTGCGGCTTCTGATGATTTGGCAATCCCAATGCGGTGCGGCA 484  
 DB 392 ACAACAGGCGCGCTGCTCATCTGCTGATTTGCGCATTCACCAACGCGGTGAGTTT 451  
 QY 485 TGGCTGCTGAGGTTCAATCTGCGCATCTCACTTTTGGGAAACCGGCGCGGAGATAGGTT 544



```

Db 452 TCGCGTCGTGGGGCTCGCTGGGGGCACTCACCGTCGCCGAGCCGGGCGCCCTGATCGGCT 511
Qy 545 TCCGCGCTCTCGCGTGTGAGTAAACCACTGGGCGATCGGCTTCCAGACGGTGTGGAGC 604
Db 512 TTTCTGGGACACAGGGGTCTATAGTGTCTCTATAGGACCCCTTCCATCCGCGCTCAAA 571
Qy 605 AGGCGGAGAATTTGTGAAAATGGTGTGATTTGATGAAATTTGTGTGGCACTCCATTTGC 664
Db 572 CGGCGGAAATCTAGGGGGGATGGGATCATGAGCGGGGTGTTCATGGAACGGGCTAC 631
Qy 665 GTGACGCGGTGGCAAAACCTCAAGTTATTGACCGGTAGAGCAACGATGCTTTT 724
Db 632 GACCGATCTGGATCGTGTGAGTGTGCTCATGAGCGCTCCCAACCGCTTCCGGGAC 691
Qy 725 CTCCAAACAATCTGTGGGTGGCA---CTTCCGATGATGAGAGCGATTCGGGCTCTCGTG 781
Db 692 CGGAGACCCCGCGCCGTAACCCGATGTGCTCCATGGAATCGGATGTGGATGCGCC 751
Qy 782 ACCCGAGAGGCTTGAATCGGGGAGATTATGAAAAGTTGGGGGAGACGCTGCAAGC 841
Db 752 GGGCGGACCGGCGGGCGTCAAGGAGCTACTGCAAGGCGCCACGAGCGGGTGTGT 811
Qy 842 TTTCTGTGCGCGTGTGCGCGCATTTGAGCCCGGCTGTGCGCGTTCCTTGGCGCGATCG 901
Db 812 TGTCAAGAAACGATCAAGGCGGAGCG---GCGACACGCTGCTGGCGCTGGCGCTTTG 868
Qy 902 GGGGCGGCGCGGTGTGCTGAATTGGGCGAG-----GATGCGCGCTTCAAGC 946
Db 869 GGGGCAACCAACGGGTGTCTCGGCGAGCAAGGCGATGAGCGCGGGGAGACACTG 928
Qy 947 TTGGGCGGAGAGCTGCGTTTGGCGCTGTGCGATTTGCTGTGCGCGCGCAAGCTAAAC 1006
Db 929 TCGGGCCGCGTGGCTTACGCGAAGCCGACGCGGGATGGCGCTGCGCCGCGATGTGTC 988
Qy 1007 TCGCGATGTGTTCATTCGACACTTCGCGCGCGCAATTGTTCGAGCGCGGTGAGAGC 1066
Db 989 TCGCGCTGTGTGTGATGACGCGCGCGGACCCGCTGTGTGCGCGCGCAAGCAAG 1048
Qy 1067 TCGGATGCGCAAGCTCGATTGGCGGCACTTGTCCAACTTATGCAAGCTCCCTCCGCA 1126
Db 1049 GGGGCTGCGCGCGAGATCGGCAATGCTGCGCGAGGCTGTCAAGCTGATACCCGCA 1108
Qy 1127 CCGTTTCGATATTATTTGGTCAAGGCGTGTGGCGTGGCGCGTGCATGCTGCCGCG 1186
Db 1109 CGGTGTGATCTGTGTGGGCGAGGCGAGGCGGCGCGCGTGGCGATGTTGCCGCG 1168
Qy 1187 ATTTGTCTAGCGGCGCGAAAGCGGTGTGTCCGATTGCGACCAAGGCGCGCTTGG 1246
Db 1169 ACCGGGTCTGCGCGCACTCCACGCGCTGGCTGGCGCCCTTGTCCGAAAGAGCCAGCG 1228
Qy 1247 CGATCTCTTCGCGGAGCAACACGCGCGGAAATCATAGAGGAGCAAGGCGGTGAGG 1306
Db 1229 CGATGTGTTCGAGACACTGTCTCATGCGCGCGAACTGCTGCCGCCCAAGGCACTCGGT 1288
Qy 1307 CGCAGCACTTTTAAAGCAAGGCTTATCGAGGAGTGTGCGCGCAAAACGAGCACTTTG 1366
Db 1289 CGGCGGACCTATGAAAGTGGGGAATTTGTGACACCATGTGTGCGGAGTAAACCCGAGCG 1348
Qy 1367 TTGAAGA 1373
Db 1349 CAGACGA 1355

```

## RESULT 6

```

US-09-815-242-7800
; Sequence 7800, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsem, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.

```

```

; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: EITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7800
; LENGTH: 873
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(873)
; US-09-815-242-7800

```

```

Query Match 8.3%; Score 122.8; DB 9; Length 873;
Best Local Similarity 52.8%; Pred. No. 1e-27;
Matches 265; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

```

```

Qy 179 ATGAATCGGTAATTAATGAGAAAGCACCGTGGAGGGCAATTCGGTATTTTGT 238
Db 326 AGAGCGGCTGATGATGCAATGAGCGGCAAGCTGAGAGGATGCGGCTGCGCCCT 385
Qy 239 CCGATTTTCTTCTTCTGCGGCGTTCCTTTGGGCAAGTGGCGGTGCGATCAATGAG 298
Db 386 TCGAGTTCTCTTCATGAGGCGGTTCGATGAGGCGCATGTCGCGAGCGCTTCTGCG 445
Qy 299 CGATTACCGGCGCACGAGCTGAAATCTCCACTGCTGTGTCTCCCTGCTCCGCTGTG 358
Db 446 CAGCAACGTGCGCTTGAAGAGCGCTGCCGCTGATCTGCTTCTCGGCTCCGCGCG 505
Qy 359 CGCGCATGCGAAGAAATCGAGCTTTTGTCAATGATGTGTCATTAACCGCGCTGTGC 418
Db 506 CGCGCATGCGAAGAAACCTCATCTCGCTGATGAGATGCGCAAGACCTGCGCGTCTGG 565
Qy 419 AGCGTACCGCGAGCGCATTTGCGTCTGTGTGATTTTGGCAATCCACATGGGTG 478
Db 566 CGCGCTGCGGAGAAAGGCAATCCGTTGCTGTGATTAACGAGCCGCTTACGGCG 625
Qy 479 GCGCATGCGCTGTGGGTTTCACTGTGGCATCTCACTTTTGGGAACCGCGCGCGAGA 538
Db 626 GCGTTTCGCGCAGCTGCGCATGCTGCGCAGCTGATCTGCGCAACCCAAAGCGCTGA 685
Qy 539 TAGGTTCTGCGGCTCTCGCGTGTGAGTTAATCACTAGGCGATGCGCTTCCAGACGCTG 598
Db 686 TCGGCTTTCGCGGCTCTCGCGTGTGAGTGAAGACCGTCCGCGAAGACTTCCGAAAGCT 745
Qy 599 TCGACGAGCGGAGAAATTTGTGAAAATGTTGTGATGATGGAATTTGTGCTCCACTTC 658
Db 746 TCGAGGCTAGAGATTCCTCTTGAGCATGCGCGCATCGACATGATGATGATGATGCGG 805
Qy 659 AATTGCTGACGCGGTGCAAA 680
Db 806 AGTTGCGGCGCGCGCTGGCCAA 827

```



```
QY 298 GGGATTACCGGCGCCACAGAGCGTAACTCCACCTGCTGCTCCCTGCTCCGGTGT 357
Db 5441 TTGTTGACGGGGCGACAGAAAAGCTTTACCAAGTATGATTTTCACTGCTGCTGT 5500
QY 358 GGGCGCATCGAGAGACAACTCGAGCTTTTGTGATGATGATGATGATGATGATGAT 417
Db 5501 GCGCGTATCGAAGAGAAATTTTTCATTGATGCAATGCGCAAAATTTTGGCGCGCTTTG 5560
QY 418 CAGGTCACCGGAGCGCATTTGCGCTTCTGCTGATTTTGGCAATCCACGATGGGT 477
Db 5561 CACCGCATAAACAAACAGGCTGCTGATCTTAAGATTAAGCTAATCAACGATGCG 5620
QY 478 GGGCGCATCGGCTCGTGGGGTTCATCTGCGCATCTCACTTTTGGCAACCGCGCGCAG 537
Db 5621 GGGTATACCGCAAGTTTGGATGATGATGATGATGATGATGATGATGATGATGAT 5680
QY 538 ATAGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 597
Db 5681 ATCGGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5740
QY 598 GTGCGACGCGCGGAGAAATTTGTAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 657
Db 5741 TTTCAAAAGGCGCAAGTTTCTTTAGAACATGCTTTTGTATGATGATGATGATGAT 5800
QY 658 CAATTGCGT 666
Db 5801 CTTTGGCT 5809
```

```
RESULT 9
US-09-815-242-7288
; Sequence 7288, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseibeck, Robert
; APPLICANT: Onisen, Karl L.
; APPLICANT: Zykith, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT FILING DATE: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7288
; LENGTH: 870
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(870)
US-09-815-242-7288
Query Match 8.1%; Score 119.2; DB 9; Length 870;
```

```
Best Local Similarity 50.9%; Pred. No. 1.3e-26;
Matches 283; Conservative 0; Mismatches 273; Indels 0; Gaps 0;
QY 93 TTCTGATGATGAATCTCCCAATATGACAACTCAATCAAGCTATGACAGACTTGA 152
Db 240 TTACGCGCTATATGATCTTTAAATTTGATGATGAAGATTAATAACACGATTA 299
QY 153 GCGGCTCGAAGCAAGCCAAATGCGATGAATCGGTAAATTACTGAGAAAGGACCGTGA 212
Db 300 AAAATACAAAAGAAAGACTAACCGCCCAAGCTGATGACAGGCTGAAGCTAAATCA 359
QY 213 GGGCATTCGGTAGCGCTTATTTTGTCCGATTTTCTTCTCTGCGGCTTTTGGGAC 272
Db 360 CCGCATGCTTTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 419
QY 273 GGTGCGTGTGCGCATGACAGCGATTAACCGCCACAGAGCTGAATCTCCACT 332
Db 420 TGTGAGGGGCAAAAGATGTAAGAGCAATCAATGCGCGGTGCTAAAGAAAGCGTT 479
QY 333 GCTGCTCTCCCTGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 392
Db 480 ATGATTTTGAAGAGTGGGGGGCTGAGTGAAGATCACTTATTCCTCATGCA 539
QY 393 GATGCTGCATTAACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 452
Db 540 AATGCTAAAGAGAGCGGCTTTGAACGATGATGAGGCAAACTCCCTTCAATTTC 599
QY 453 GATTTTCCCAATCCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 512
Db 600 GCTCTTAAGCGATCCCATTTATGAGGGGCTTATGCGCATTTTGGCTTTTATGAGGATCT 659
QY 513 CACTTTGGGAAACCGGCGCGCATAGATGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 572
Db 660 CATTATCGAGAGCCAGGGCGCATGATGCTTTGGGGGCTTATGAGGATTAAGCAAA 719
QY 573 CACTGCGCATGCGCTTCCAGAGCTGTGACAGCGCGGAGATTTGTAAAACTGTGT 632
Db 720 TATAGGGGCGGATTTGCTGAGGGCTTTCAACAGCGGAATTTTATAGAGATGCTT 779
QY 633 GATTGATGATGCTG 648
Db 780 GATTGATGATGCTG 795
```

```
RESULT 10
US-09-895-913A-97
; Sequence 97, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleinhous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Cohen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; FILE REFERENCE: 06132/043002
; CURRENT FILING DATE: US/09/895,913A
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 06/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 1053
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (110)...(976)
US-09-895-913A-97
```

Query Match 8.1%; Score 119.2; DB 10; Length 1053;  
 Best Local Similarity 50.9%; Pred. No. 1.4e-26;  
 Matches 283; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

QY 93 TTCTGGATGMAACTCCCAATATACAACTCATCAAGGTATGCGAGACCTTGG 152  
 DB 349 TTTAGCGCTTAATGATCTTTTAATTTCTGATTAAGAGAGCTATTAACAACGCAATT 408  
 QY 153 GCGGGCTCAGCAAGGCCAAATGCGATGAATCGTAATTAATTAATGAGAGGACCGTGA 212  
 DB 409 AAAATACGAAAAAGACTTAACCGCCCAAGCTCAAGTACAGGGGTGAAGCTTAATCA 468  
 QY 213 GGGCAATCCGGTAGCCGTTATTTTGTCCGATTTTCTTCTCGCGGCTTCTTGGGCA 272  
 DB 469 CCGCATGCTTTGACAGATGAGTGTGTTGATTTTACCTTTAATGAGGGGAGTTAGGCT 528  
 QY 273 GGTGCGGTGCGGATGATGAGGCAATTCACCGCCGACAGAGCTGAACTCCACT 332  
 DB 529 TGTGAGGGGCGAAGATGTAAGACATCAATCCGGGCTGCTMAAAGAGAGCGTT 588  
 QY 333 GCTGCTCTCCCTGCTTCCGCTGCTGCGGCAATGCAAGCAATCGAGCTTTGTCA 392  
 DB 589 ATTTGATTTTTCAGCAAGTGGGGGCTTGATGCAAGATCACTTATTCGCTATGCA 648  
 QY 393 GATGCTGTCATPACCGCGCTGTGCAAGCTCAACGAGGCGCATTTGCGTCTGCT 452  
 DB 649 AATGCTMAAACGAGCGGCTTTGAACGATTAAGTGAAGGCAAACTCCCTTATTTTC 708  
 QY 453 GATTTTGGCAATCCCAAGATGGGTGGCCATGCTGCTGAGGTTATCTGGGCAAT 512  
 DB 709 GCTTTAAGCGATCCCACTTAATGAGGGGCGTTAGCGGATCTTTTGTGAGGATCT 768  
 QY 513 CACTTTTGGGGAACCGGCGGCGAGTAGTTTCTCGGCTCCGCGGTGGTGAAGTTAC 572  
 DB 769 CATTTTCGAGAGCCAGGGGCGCATTAATAGGCTTTGGGGGCTTAGGGTATTAAGCAAG 828  
 QY 573 CACTGGGCAATGCTTCCAGAGCGTGTGCAAGGCGGAGAAATTTGTGAAAACTGTGT 632  
 DB 829 TATAGGGGCGGATTTGCTGAGGGCTTCAACAGCGGAATTTTATTAAGACATGGCT 888  
 QY 633 GATTGATGAATTTG 648  
 DB 889 GATTGATGATTTG 904

RESULT 11  
 US-09-815-242-9854  
 ; Sequence 9854, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; FILE REFERENCE: ELITRA.011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; PRIOR FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 9854  
 ; LENGTH: 993  
 ; TYPE: DNA  
 ; ORGANISM: Salmonella typhi  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)...(993)  
 US-09-815-242-9854

Query Match 7.3%; Score 108; DB 9; Length 993;  
 Best Local Similarity 49.6%; Pred. No. 4.5e-23;  
 Matches 276; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

QY 131 AAGGCTATGAGAGCTTTGAGCGGGCTCGAAGCAAGGCCAAATGCAATGATGAT 190  
 DB 344 AAAATATTAAGACAGACTGGCTGCGCGCAAGAAACCGCGAAGACGCGCTGG 403  
 QY 191 TTACTGAGAAAGCACCGTGAAGGGCATTCGGTAGCCGTTATTTTGTCCGATTTTCT 250  
 DB 404 TGTGATGAAGAGGACGCTTACAGGTATGCGGTTGTGCGCGCGCTTGAATTCGCT 463  
 QY 251 TCTTGGCGGTTCTTTGGGCAAGTGGCTGCTGCGTGCATATGAAGCGATTCACGCG 310  
 DB 464 TCATGGGCGGCTCATAGGGGCTCTGCTTGGCGCAAGCTTGTGCGCGCTTGAACAG 523  
 QY 311 CCAGAGCTGAAATCCCACTGCTGCTCCCTGCTTCCGCTGCTGCGGCAATGCAAG 370  
 DB 524 GCTGGAAGACACTGCTGTTAGTGTCTTCTGCTTCCGCGCGCGGTATGCAAG 583  
 QY 371 AAGCAATGAGCTTTTGTATGATGATGTCATTAACCGCGGCTGTGAGCGTCAACGCG 430  
 DB 584 AAGCGCTATGCTGATGATGAGATGGGAAACCTGCGCGCGCTGCTAAATGCAAG 643  
 QY 431 AAGCGCATTTGCGCTCTGCTGATTTTGGCAATCCAGATGGTGGCCCATGCT 490  
 DB 644 AACGTGCTGCTCCCTACATCTCGGATTAAGCCGATCCGACATGGGCGGCTTCCGCA 703  
 QY 491 CGTGGGTTTATCTGGGATTTCACTTTTGGGAACCGGGCGGCAATGATTTCTCG 550  
 DB 704 GTTTGCGATCTGGGCGATCTCAACATGCGCGAAGCCCTGATTTGCTTCCGCG 763  
 QY 551 GTCTCGGCTGTGAGTTAACCACTGGGCAATGCGCTTCCAGACGTTGCGAGCGG 610  
 DB 764 GCCCGCGGTTATGCAACAAACGTTGCTGAGAAAGCTCCCGCAGATTTCAAGCGCAT 823  
 QY 611 AGAATTTGTGAATACTGTGTGATTAATGGAATTTGTGCGCACTCAATTTGCTGAG 670  
 DB 824 AGTTCTGATGAAAGAGCGCTATTGATATGATGTCGCGCGCGGAATGCGCGCTGA 883  
 QY 671 CGGTGCAAAAACCT 686  
 DB 884 AGTGGCGAGCATTTCT 899

RESULT 12  
 US-09-815-242-6133  
 ; Sequence 6133, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23,207,727  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6133  
LENGTH: 915  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(915)  
US-09-815-242-6133

Query Match 7.0%; Score 102.8; DB 9; Length 915;  
Best Local Similarity 48.9%; Pred. No. 1.8e-21;  
Matches 275; Conservative 0; Mismatches 287; Indels 0; Gaps 0;

200 AAGCACCCTGAGAGGAGGATCCGCTGATGCTTATTTTCTCCGATTTTCTTCTCCGCG 259  
335 AAGCACCCTGATGAGATGCGGCTGCTGCGGATTCGAGTTCCTTATGCGCG 394  
260 GTTCTTTGGGACCGTCCGCTGCGGATTCGAGTTCGAGTTCCTTATGCGCG 319  
395 GTTCAATGGGCTGCTGCTGCGGATTCGAGTTCGAGTTCCTTATGCGCG 454  
320 TGAATCTCCACTGCTGCTCCCTGCTCCGCTGCGGATTCGAGTTCCTTATGCGCG 379  
455 ATAACTGCGGCTGCTGCTCCCTGCTCCGCTGCGGATTCGAGTTCCTTATGCGCG 514  
380 GAGCTTTTGTATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 439  
515 TGTGCTGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 574  
440 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 499  
575 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 634  
500 CATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 559  
635 TGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 694  
560 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 619  
695 TTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 754  
620 TGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 679  
755 TCGGAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 814  
680 AAACCTCAAGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 739  
815 GCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 874  
740 GCGTGGCACTTCCGCTGATGA 761  
875 TAGTGTATCCCCCGTATCCGGA 896

RESULT 13  
US-09-815-242-9520  
Sequence 9520, Application US/09815242  
Patent No. US20020061569A1

GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.

TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9520  
LENGTH: 867  
TYPE: DNA  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(867)  
US-09-815-242-9520

Query Match 6.8%; Score 97.8; DB 9; Length 867;  
Best Local Similarity 49.7%; Pred. No. 6.6e-20;  
Matches 249; Conservative 0; Mismatches 252; Indels 0; Gaps 0;

179 ATGAATCGTAACTACTGAGAAAGGACCGTGGAGGATTCGCTGATGCTTATTTGT 238  
332 ATGAAGCGTTTGACAGAAAGCTCTTATTAAGGTGAGACTGTGGATTA 391  
239 CCGATTTTCTTCTTCCGCTGCTTCTTGGGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 298  
392 TGATTTCTAATCTTATCATGCTCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 451  
299 CGATTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 358  
452 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 511  
359 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 418  
512 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 571  
419 AGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 478  
572 AAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 631  
479 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 538  
632 GTGTGAGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 691

QY 539 TAGTTTCTGGGTCCTGGCGTGTGGAGTTAACCACTGGGCGATGCGCTTCAGACGGTG 598  
DB 632 TTGGTTTGTGGGGCGTGGTGTGATTTGAATAACGGTGTGTAAGCTTGCGTAGGAT 751  
QY 599 TGCACAGCGGAGAAATTTGGTGAACCTGGGTGTGATTTGATGAAATTTGTGCGCACTCC 658  
DB 752 TCCAAAGGCGAAATTTCTATTAGAACATGAGCTTTGTGATGCTATTGTCAAAAGAGAG 811  
QY 659 AATGCGTGCAGCGGTGGCAA 679  
DB 812 ACTTACGAGATACGATTGCTA 832

RESULT 14  
US-09-974-300-5531  
Sequence 5531, Application US/09974300  
Patent No. US20020146721A1  
GENERAL INFORMATION:  
APPLICANT: Beirka, Randy M.  
APPLICANT: Clausen, Ib Groth  
TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
FILE REFERENCE: 10085-500-US  
CURRENT FILING DATE: US/09/974.300  
PRIOR FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: 09/680,598  
PRIOR FILING DATE: 2000-10-06  
PRIOR APPLICATION NUMBER: 60/279,526  
NUMBER OF SEQ ID NOS: 8481  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5531  
LENGTH: 654  
TYPE: DNA  
ORGANISM: Bacillus clausii  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(654)  
OTHER INFORMATION: n = A,T,C or G  
US-09-974-300-5531

Query Match 6.6%; Score 96.8; DB 10; Length 654;  
Best Local Similarity 51.8%; Pred. No. 1.2e-19;  
Matches 288; Conservative 0; Mismatches 264; Indels 4; Gaps 3;  
QY 129 TCAAGCGTATGACAGACTTGGAGCGGCTTGAAGCAAGGCCAAATGCGATGATCGGT 188  
DB 84 TCCAGTTATAGAGAAAGCTTGAGAGCGACCGTAAACCCGTTGAAATGAAGCGGT 143  
QY 189 AATTACTGAGAGGACCGTGGAGGCGATTCGGTAGCCGTT-ATTTGTCGATTTT 247  
DB 144 GGTCACTGTGAAGAAAGATTAATGATATCCGTTGATTCATTCAGATTATGAGCTCA 203  
QY 248 CTTTCTCGGCGGCTTCTTTGGGACGCTCGGCTCGGTGCGATCATGAAGCGATTCACC 307  
DB 204 ATTTTCGATGCGAGCATGGGCTCGGCGTTGGGAAAGCTAACAGGGCAATTGAT 263  
QY 308 GCGCCACAGACTGAAATCTCCACTGCTGG--TCTCCCTGCTTCGGGTGTGCGCGAT 365  
DB 264 TGGGCACCTGAGACGAGCGCTTCGCTTNAATTTTGGCTGCTCAGGGGAGCGCGCAT 323  
QY 366 GCAGGAGACATGAGCTTTTGTGATGATGTGTCATAACCG-CGGCTGTGACGCTC 424  
DB 324 GCAAGAGGAGCATGCTCAGACCTTAATGCAATGCGAAACGAGCGCGCANTTGAAGCTTC 383  
QY 425 ACCGCGAGCGCATTTGCGCTTCTGTATATTTGCGCAATCCACGATGGGTGGCGCA 484  
DB 384 TGAACCGAGCAGAGGGCTTACATTTGCGTTATGACCAACCGACGAGCGGCGCGGT 443  
QY 485 TGGCTCTGTGGGTTCACTGTGGGATCTCACTTTTGGGAAACCGGCGCGAGATAGTT 544  
DB 444 CCGGATGTTTGGCGCTCTCTGGCGATTACAATTTTGGCGAGCCAAAMGCACTGATTGCT 503

QY 545 TCTGGGTCTTCGCGGTGGAGTTAACCACTGGGCGATGCGCTTCAGACGGTGTGACG 604  
DB 504 TCGCTGGAGGCGCATTTTGAACGACCATTCGTCAGGAACCTTCTGAAGCTTCAGA 563  
QY 605 AGCGGAGAAATTTGGTGAACCTGGGTGTGATTTGATGAATGTGTGCGCACTCCAATTGC 664  
DB 564 CAGCGAAATTTCTTTGAGAGATGGCCAGCTTGATGATGATTCATTCAGAGTTAAGATGA 623  
QY 665 GTGCAACCGGTGGCAA 680  
DB 624 AAGAAACGCTGACTTA 639

RESULT 15  
US-09-815-242-7115  
Sequence 7115, Application US/09815242  
Patent No. US2002061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zykkind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815.242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7115  
LENGTH: 891  
TYPE: DNA  
ORGANISM: Haemophilus influenzae  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(891)  
US-09-815-242-7115

Query Match 6.5%; Score 96.2; DB 9; Length 891;  
Best Local Similarity 49.5%; Pred. No. 2.1e-19;  
Matches 248; Conservative 0; Mismatches 253; Indels 0; Gaps 0;  
QY 179 ATGAATGGTAAATTAAGTGAAGCAACCGTGGAGGCGATTCGGTATGCGCTTATTTGT 238  
DB 320 AAGATGGCTAATTAATTAAGTGAAGTACATTTTAATATGCAATGCTTTGGCTGAT 379  
QY 239 CCGATTTTCTTCTCGGCGGCTTCTTTGGGACGCTCGGTCGGTGCATCATGAAG 298  
DB 380 CGAATTTGCTTTATGAGGCGGCTTCAATGGGTTCTGTATGATGGTGAATTTGTTAAG 439  
QY 299 CGATTACCGGCGACAGAGCTGAATCTCCACTGCTGTCTCCCTGCTTCGGTGTG 358  
DB 440 CCGCTGAAGAAAGCGATGAATGAATGTCCATTTGTGTGTTCTTCGCGAGTGTGTG 499  
QY 359 CGCGCATGAGAGAAATCATGAGCTTTTGTATGATGATGTGTCCATTAACCGCGCTGTC 418

Db	500	CTCGTATGCAGAGACATTATTCTTTAATGCAAAATGCAAAAATAGTGCCTGACTTG	559
Qy	419	AGCGTCAACCGGAGAGCGCATTTGCGCTCCGCTGTATTTGGCAATCCCAAGATGGTG	478
Db	560	CTCAATGCGTGAAGGGGTGTGCCATTATTTCAGATTAACGGAATCGACTTTAGGCG	619
Qy	479	GCGCCATGCGCTCGTGGGGTTCATCTGGGCATCTCACTTTTGGGAACCGGCGGCAGA	538
Db	620	GCGTATCAGCCAGTTTTCGATGTTAGGGGATTTAATATTCGAGCCAAAAGCCTTA	679
Qy	539	TAGGTTCTGGGTCTCGCTGTGAGTTAACCATGCGCATGCGCTTCCAGACGGTG	598
Db	680	TTGGTTTGCAGGCGCAGCGTTAATGAACAACCTGCGTGAATAAATTGCCAGAAAGGTT	739
Qy	599	TGCAGCAGGCGGAGATTGTGTAATACTGGTGTGATGATGAATTGTGCGCCTCC	658
Db	740	TCCAACGTAGTAGTTCTTACTTGAGAAAGGGCAATTGATATGATCGTGAACGTTTCAG	799
Qy	659	AATGCGTGCAGCGGTGSCAA	679
Db	800	AAATCGTCAAACTTTAGCAA	820

Search completed: November 13, 2003, 06:16:54  
 Job time : 912.783 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 00:15:23 ; Search time 75 Seconds

(without alignments)  
1039.130 Million cell updates/sec

Title: US-10-024-370-3

Perfect score: 2448

Sequence: 1 VEKPEPTVMWGMEHTSALT.....NNPERAGRDSRFRPERLQAQ 491

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A.GeneSeq\_15Jun03.\*  
1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2448	100.0	491	22	AA045789
2	2445	99.9	491	22	AA090675
3	2411	98.5	484	22	AA076522
4	1047	42.8	495	22	AA081130
5	400	16.3	288	22	AA035332
6	384	15.7	289	19	AA098434
7	384	15.7	289	22	AA035792
8	383	15.6	225	23	AA051006
9	375	15.3	330	22	AA038359

10	374	15.3	304	22	AA034637
11	368	15.0	288	22	AB054092
12	368	15.0	296	22	AA035619
13	365	14.9	291	23	AB028019
14	364	14.9	312	20	AA036893
15	362	14.8	279	23	AB050977
16	360.5	14.7	1106	22	AB025523
17	354.5	14.5	285	22	AA083049
18	354.5	14.5	293	23	AB038958
19	344.5	14.1	292	24	AB078030
20	343.5	14.0	285	22	AA037025
21	343.5	14.0	285	22	AA073268
22	340.5	13.9	252	22	AA033928
23	335	13.7	288	23	AB028020
24	334	13.6	288	22	AA037772
25	334	13.6	288	22	AA038025
26	334	13.6	288	22	AA030264
27	334	13.6	288	24	AA0000780
28	329.5	13.5	290	22	AA036305
29	319.5	13.1	255	20	AA034658
30	318.5	13.0	294	23	AB048048
31	307	12.5	527	22	AA096609
32	300.5	12.3	318	23	AB048047
33	291.5	11.9	256	23	AB028022
34	291.5	11.9	280	22	AA083050
35	289	11.8	317	18	AA027453
36	288	11.8	315	22	AA035421
37	288	11.8	315	24	AA030458
38	285.5	11.7	540	23	AB06186
39	285	11.6	527	23	AB098725
40	283.5	11.6	315	23	AB038941
41	282	11.5	875	19	AA048309
42	277	11.3	309	22	AA034396
43	275	11.2	314	22	AA037315
44	275	11.2	526	24	AB057631
45	274	11.2	314	22	AA072365

#### ALIGNMENTS

RESULT 1  
ID AAB45789  
AAB45789 standard; Protein; 491 AA.  
AC AAB45789;  
XX  
DT 15-MAR-2001 (first entry)  
XX  
DE C. glutamicum accDA protein.  
XX  
KW L-amino acid; accDA; coryneform microorganism; L-lysine; animal feed; medicine; pharmaceutical industry.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN EPI055725-A2.  
XX  
PD 29-NOV-2000.  
XX  
PF 10-MAY-2000; 2000EP-0109842.  
XX  
PR 27-MAY-1999; 99DE-1024365.  
XX  
XX (DEGS) DEGUSA-HUELS AG.  
PA (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.  
XX  
PI Tilg Y, Eggeling L, Birkmann B, Sahm H, Moeckel B;  
DR WPI; 2001-042411/06.  
XX N-PSDB; AAC82732; AAC82733.  
PT Cloned Corynebacterium glutamicum accDA gene useful for producing

transformed coryneform bacteria producing increased yields of L-amino acids, especially L-lysine -

Claim 3; Page 15-17; 20pp; German.

This invention describes a novel cloned Corynebacterium glutamicum DNA (1), replicable in coryneform microorganisms coding for an accD gene. The invention also describes (1) a polypeptide derived from (1) having a fully defined 491 aa sequence; (2) coryneform microorganisms transformed with one or more copies of (1); (3) the shuttle vector pZiacCDa contained in Corynebacterium glutamicum DSM 12785; and (4) a process for producing L-amino acids, comprising culturing a coryneform bacterium that overexpresses the accD gene. Coryneform bacteria transformed with (1) so that they overexpress the accD gene are useful for producing L-amino acids, especially L-lysine, which are useful in animal feeds, in human medicine and in the pharmaceutical industry.

Sequence 491 AA:

Query Match 100.0%; Score 2448; DB 22; Length 491;

Best Local Similarity 100.0%; Pred. No. 7.66-215; Mismatches 0; Indels 0; Gaps 0;

Matches 491; Conservative 0;

1 VEKRPPTVMWGMETSALTLLIDSVLPDPSFISWNETPOYDNLNGVAETLERARSKACD 60  
1 VEKRPPTVMWGMETSALTLLIDSVLPDPSFISWNETPOYDNLNGVAETLERARSKACD 60

61 ESVITGEGTVEGIPVAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGGA 120  
61 ESVITGEGTVEGIPVAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGGA 120

121 RMOEDNRAFWMMVSTTAAVQHRRAHLPFLVYLNPMTMGAMASWSSGHLTFEPGQOI 180  
121 RMOEDNRAFWMMVSTTAAVQHRRAHLPFLVYLNPMTMGAMASWSSGHLTFEPGQOI 180

181 GFLGRVVELTTGHALPDGVQQAENLVKTVGDVSPLOQRAAVAKTLKVQPEATDR 240  
181 GFLGRVVELTTGHALPDGVQQAENLVKTVGDVSPLOQRAAVAKTLKVQPEATDR 240

241 FSPPTPGVALPVMEAIARSRDPQRPQIGIEMETLGADVVKLSGARAGALSPAVRALARI 300  
241 FSPPTPGVALPVMEAIARSRDPQRPQIGIEMETLGADVVKLSGARAGALSPAVRALARI 300

301 GGRPVVLIGDPRRTFLGPOELRFARRGISLAREINLPVSIIDTSGAELSGAAEELGIAS 360  
301 GGRPVVLIGDPRRTFLGPOELRFARRGISLAREINLPVSIIDTSGAELSGAAEELGIAS 360

361 STARTLSKLIDAPLPTVSITIGGVGGGALAMLPADLVYAENAMLSALPEGASAILFR 420  
361 STARTLSKLIDAPLPTVSITIGGVGGGALAMLPADLVYAENAMLSALPEGASAILFR 420

421 DTNHAABEIIERQVAAHALISQGLIDGIVAEETHEFVEEILGTISNALSELNNPERAGRD 480  
421 DTNHAABEIIERQVAAHALISQGLIDGIVAEETHEFVEEILGTISNALSELNNPERAGRD 480

481 SRFTRPERIAQ 491  
481 SRFTRPERIAQ 491

RESULT 2  
AAG90675 ID AAG90675 standard; Protein; 491 AA.

AAG90675;  
XX

26-SEP-2001 (first entry)

C glutamicum protein fragment SEQ ID NO: 4429.

Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

organic acid synthesis.

OS Corynebacterium glutamicum.

EP1108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-0127688.

16-DEC-1999; 99JP-0377484.

07-APR-2000; 2000JP-0159162.

03-AUG-2000; 2000JP-0280988.

(KYOW) KYOWA HAKKO KOSYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

Tateishi N, Senoh A, Ikeda M, Ozaki A;

WPI; 2001-376931/40.

N-PSDB; AAH65894.

Novel polynucleotides derived from Coryneform bacteria, for identifying

mutation point of a gene, measuring expression of a gene, analysing

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

expression profile or pattern of a gene and identifying homologous gene

```

DB 361 STARTLSKIDLPFTSVIIGGVGGALAMLPADVVAANAMLSALPFGASAILFR 420
QY 421 DNNHAEIIRERGVQAHALLSGGLIDGVAETHEHVEEILGTISNALSELNNPERAGRD 480
DB 421 DNNHAEIIRERGVQAHALLSGGLIDGVAETHEHVEEILGTISNALSELNNPERAGRD 480
QY 481 SRTFRERLAQ 491
DB 481 SRTFRERLAQ 491

RESULT 3
AAB76522
ID AAB76522 standard; Protein; 484 AA.
XX
AC AAB76522;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein SHQ ID NO:26.
XX
KW Corynebacterium glutamicum; brevivibacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering.
XX
OS Corynebacterium glutamicum.
XX
PV MO200100805-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00926.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.
PR 27-AUG-1999; 99DE-1040833.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041395.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042088.
XX
XX (BADI ) BASF AG.

```

```

XX XX
PI Pompejus M, Kroege B, Schroeder H, Zelder O, Haberhauser G;
XX DR MPI: 2001-071486/08.
XX DR N-PSDB; AAF67755.
XX
PT Corynebacterium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
PT transformation.
XX
PS Claim 20; Page 175-177; 1119pp; English.
XX
CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC construction and membrane transport (MCT) proteins given in AAB76510 to
CC AAB76847. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevivibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention.
XX
SQ Sequence 484 AA;
XX
Query Match 98.5%; Score 2411; DB 22; Length 484;
Best Local Similarity 100.0%; Pred. No. 1.0e-211;
Matches 484; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 MWGMHESATLTLDVSDPDSFISWNETPOYDNLNCGAETLERARSKACDESVTNGE 67
DB 1 MWGMHESATLTLDVSDPDSFISWNETPOYDNLNCGAETLERARSKACDESVTNGE 60
QY 68 GTVEGIPVAVILSDPSFLGSLGTVASVRIKAIHRATELKLPLVSPASGARMQEDNR 127
DB 61 GTVEGIPVAVILSDPSFLGSLGTVASVRIKAIHRATELKLPLVSPASGARMQEDNR 120
QY 128 AFVWVNSTTAQVQRREAHLPFLVYLRRPTMGAMASVSGSHLTFARPQAQIGLGRV 187
DB 121 AFVWVNSTTAQVQRREAHLPFLVYLRRPTMGAMASVSGSHLTFARPQAQIGLGRV 180
QY 188 VELTTGHALPDGVQAAEHLVKTGVYIDGIVSPLOJRAAVALTKYIQPEATDRFSPPTPG 247
DB 181 VELTTGHALPDGVQAAEHLVKTGVYIDGIVSPLOJRAAVALTKYIQPEATDRFSPPTPG 240
QY 248 VALPMEAIARSRDPORPGIEIMETLGADVVKLSGARAGALSPAVRVALARIGRPPVL 307
DB 241 VALPMEAIARSRDPORPGIEIMETLGADVVKLSGARAGALSPAVRVALARIGRPPVL 300
QY 308 IGDRRFTLGPQELRPARGISLARELNLPIVSIITDTSGAEISQAEEGLIASSIARTLS 367
DB 301 IGDRRFTLGPQELRPARGISLARELNLPIVSIITDTSGAEISQAEEGLIASSIARTLS 360
QY 368 KLIDAPLPTVSVIIGGVGGALAMLPADVVAANAMLSALPFGASAILFRDNNHAE 427
DB 361 KLIDAPLPTVSVIIGGVGGALAMLPADVVAANAMLSALPFGASAILFRDNNHAE 420
QY 428 IIRGVQAHALLSGGLIDGVAETHEHVEEILGTISNALSELNNPERAGRDSEFTFE 487
DB 421 IIRGVQAHALLSGGLIDGVAETHEHVEEILGTISNALSELNNPERAGRDSEFTFE 480
QY 488 RLAAQ 491
DB 481 RLAAQ 484

RESULT 4
AAG81130
ID AAG81130 standard; Protein; 495 AA.
XX
AC AAG81130;
XX

```

[illegible]

Db	306	GGSTVGPAAALFEARKRGMAALAECLPVLTVIDAAGPALSAAEQGGALAGJAHCLAEIVT	365
Qy	372	APLPFTVSVITIQGVGGGALAMLPADLYVAAENAMLSALPEGASAILPRDTNHAETIER	431
Db	366	LDTPFTVSTILCOGSGGAPALAMLPADRVLAALHGMALAPPEGASAIYFRDTHAAELAAA	425
Qy	432	QGVVAHALLSQGLLDGIVAEETEHFVEETIG---TTSNALSELDN-----NPERAGRDSR	482
Db	426	QGRNADLLKSGKIVDTIVPEYIPDAADEFTFPAALSNIAAEVHAIRKTIPIAPERLA--TR	483
Qy	483	FTFPERL 489	
Db	484	LQRYRRI 490	
RESULT 5			
AAU35332	ID	AAU35332 standard; Protein; 288 AA.	
XX	AAU35332;		
DT	14-FEB-2002	(first entry)	
XX	Enterococcus faecalis	cellular proliferation protein #619.	
XX	Antisense;	prokaryotic cellular proliferation protein;	
KW	antibiotic;	antibacterial; drug design.	
XX	Enterococcus faecalis.		
PN	WO200170955-A2.		
XX	27-SEP-2001.		
PF	21-MAR-2001;	2001WO-US09180.	
XX	21-MAR-2000;	2000US-191078P.	
PR	23-MAY-2000;	2000US-206848P.	
PR	26-MAY-2000;	2000US-207127P.	
PR	23-OCT-2000;	2000US-242578P.	
PR	27-NOV-2000;	2000US-253625P.	
PR	22-DEC-2000;	2000US-257931P.	
PR	16-FEB-2001;	2001US-269308P.	
XX	(ELIT-)	ELITRA PHARM INC.	
PI	Haselbeck R,	Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;	
PI	Yamamoto RT,	Xu HH;	
XX	WPI,	2001-611495/70.	
DR	N-PSDB;	AAS53191.	
XX	New polymucleotides	for the identification and development of	
PT	antibiotics,	comprise sequences of antisense nucleic acids -	
XX	Example 3;	Seq ID No 10925; 511p; English.	
XX	The invention	relates to antisense inhibitors of genes essential to	
CC	prokaryotic	cellular proliferation, their use in identifying the	
CC	genes, their	use in the discovery of novel antibiotics, the essential	
CC	genes themselves	and the encoded proteins. The prokaryotes used are	
CC	Escherichia coli,	Staphylococcus aureus, Salmonella typhi, Klebsiella	
CC	pneumoniae,	Pseudomonas aeruginosa and Enterococcus faecalis. The	
CC	invention is	also useful for the identification of potential new targets	
CC	for antibiotic	development. The antisense nucleic acids can also be used	
CC	to identify	proteins used in proliferation, to express these proteins,	
CC	and to obtain	antibodies capable of binding to the expressed proteins.	
CC	The proteins	can be used to screen compounds in rational drug discovery	
CC	programmes. The	antisense nucleic acid sequence is also useful to screen	
CC	for homologous	nucleic acids which are required for cell proliferation in	
CC	a wide variety	of organisms. The present sequence represents an	
CC	essential	prokaryotic cellular proliferation protein.	

CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences.

XX Sequence 288 AA;

Query Match 16.3%; Score 400; DB 22; Length 288;  
Best Local Similarity 42.6%; Pred. No. 5.2e-28;  
Matches 89; Conservative 39; Mismatches 79; Indels 2; Gaps 2;

QY 25 LDPOSFISM-NETPOYDIN-OGVAETLERARSKAKDESVITGEGTGVAVILSDP 82  
DB 73 VDEKSFHMDSELTAVKDPINFPQYIEKIEKQEKGLDEAVLTGKATIEGQAVAIQINDA 132  
QY 83 SFLGSLGTVASVRIMKAIHRATELKLPLVSPASGARMQEDNRAFYMMVSTPAVQHR 142  
DB 133 NFIMSGMGTIVGEKTRFERATEKGLPVIFTPASGARMQEGIFSLMQMAKISALQRH 192  
QY 143 REAHLPFLVLRNPTMGAMASWGSSGHLTPAEPGAQIGFLGPRVVELTTGHALPDGVOQ 202  
DB 193 NKAGLLVLTVDLPFTTGGVTASFAMDGDIILAEPSGLIGFAGRRVIEOTIRQELPDDFOK 252  
QY 203 AENLVKTVIGDIYSPLQIRAAVAKTLKV 231  
DB 253 AEFLIEHGFVDQIVPRNLLRQLRSLDLRL 281

RESULT 6  
AAW98434  
ID AAW98434 standard; Protein; 289 AA.

XX AAW98434;

XX 31-MAR-1999 (first entry)

XX H. pylori GHP0 451 protein.

XX GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;  
XX peptic ulcer disease.

XX Helicobacter pylori.

XX WO9843478-A1.

XX 08-OCT-1998.

XX 01-APR-1998; 98WO-US06371.

XX 29-JUL-1997; 97US-0902615.

XX 01-APR-1997; 97US-0833457.

XX 24-JUN-1997; 97US-0881227.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.

XX Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;

XX WPI; 1998-542293/46.

XX N-PSDB; AAX14153.

XX New isolated Helicobacter polynucleotides - used to develop products

XX for the diagnosis, prevention and treatment of Helicobacter

XX infections and gastrointestinal diseases

XX Claim 8; Page 738-739; 2054pp; English.

XX This sequence represents a Helicobacter pylori GHP0 protein of the

XX invention. The polypeptides can be used for preventing or treating

XX Helicobacter infections, and gastroduodenal diseases associated with

XX these infections, including acute, chronic, and atrophic gastritis, and

XX peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be

XX used for the production of antibodies. The products can also be used for

CC detection and diagnosis.

XX Sequence 289 AA;

Query Match 15.7%; Score 384; DB 19; Length 289;  
Best Local Similarity 39.3%; Pred. No. 1.5e-26;  
Matches 84; Conservative 40; Mismatches 72; Indels 18; Gaps 2;

QY 26 DPDSFISNNEPQYDNLNQGVAETLERARSKAKDESVITGEGTGVAVILSDPFL 85  
DB 85 DPLNFVD-----KESYKORIKKYEKNTNPPSSVISEAKINRMPLOIVPDFSFM 134  
QY 86 GSGISGTVASVRIMKAIHRATELKLPLVSPASGARMQEDNRAFYMMVSTPAVQHR 145  
DB 135 GSGISGTVASVRIMKAIHRATELKLPLVSPASGARMQEDNRAFYMMVSTPAVQHR 194  
QY 146 HLPFLVLRNPTMGAMASWGSSGHLTPAEPGAQIGFLGPRVVELTTGHALPDGVOQ 205  
DB 195 KLPFLISLSDPFTTGGVTASFAFLDGLIIEFGAMIGFAGPRVIVKQTIGADLPFGQTAEF 254  
QY 206 LVKTVIGDIYSPLQIRAAVAKTLKV 231  
DB 255 LIEHGLIDMIVHRDCLKTSLDLIAMMTHKTSKI 288

RESULT 7  
AAU35792  
ID AAU35792 standard; Protein; 289 AA.

XX AAU35792;

XX 14-FEB-2002 (first entry)

XX Helicobacter pylori cellular proliferation protein #105.

XX Antisense; prokaryotic cellular proliferation protein;

XX antibiotic; antibacterial; drug design.

XX Helicobacter pylori.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207727P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253625P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlson KU, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX N-PSDB; AAS53651.

XX New polynucleotides for the identification and development of

XX antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 11385; 51pp; English.

XX The invention relates to antisense inhibitors of genes essential to

XX prokaryotic cellular proliferation, their use in identifying the

XX genes, their use in the discovery of novel antibiotics, the essential

XX genes themselves and the encoded proteins. The prokaryotes used are

XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella

XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The







XX FR2807446-A1.  
 XX 12-OCT-2001.  
 XX 11-APR-2000; 2000FR-0004630.  
 XX 11-APR-2000; 2000FR-0004630.  
 XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
 XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
 XX WPI; 2002-043418/06.  
 XX New nucleotide sequence useful in the identification of Lactococcus  
 XX lactis and related species -  
 XX Claim 6; SEQ ID No 794; 2504bp; French.  
 XX The present invention is related to a Lactococcus lactis nucleotide  
 XX sequence (ABA50521) and related proteins (ABBS3300-ABBS5621). The  
 XX nucleic acid sequence is useful in the detection and/or amplification of  
 XX nucleic acid sequence, particularly to identify Lactococcus lactis or  
 XX related species. The proteins of the invention are useful for the  
 XX biosynthesis or biodegradation of a composition of interest. The  
 XX invention helps research in lactic bacteria, particularly useful in the  
 XX production of yogurt and cheese.  
 XX Note: The sequence data for this patent is based on equivalent patent  
 XX WO200177334 (published 18-OCT-2001) which is available in electronic  
 XX format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 288 AA;

Query Match 15.0%; Score 368; DB 23; Length 288;  
 Best Local Similarity 40.0%; Pred. No. 4.4e-25;  
 Matches 86; Conservative 37; Mismatches 90; Indels 2; Gaps 2;

QY 24 VLDPSFISMNETPOVDN-LN-QGYAETLERARSKACDESVITGEGTVEGIPVAVILSD 81  
 DB 72 VADKDSFVMEFTGIESKNPLDPGPEKLAATKARTGLDEAVITGTATKGGKTLATMD 131  
 QY 82 FSLGSLGTASVIRIMKAIHRAATELKLPLVSPASGARMOEDNRAFMVMTSITAAVOR 141  
 DB 132 STFIWASMGTVGKXKTRLEFVATTEKLPITIVFASGARMGEGIMSLMQAKTSAAYKR 191  
 QY 142 HREAHLPFLVLYLNPTMGAMASWSSGHLTPAEPQAQIGFLGPRVVELTTGHALPDGVQ 201  
 DB 192 HSNAGLFYITVLTDPPTGGVTASFSALGDIILAEPSLIGFAGRRVIEQTVAGTLPPDFQ 251  
 QY 202 QAEMLVKTGVIDGIVSPDLQRAAVAKTLKVIQPE 236  
 DB 252 KAEFLINHGFDVAIVKRTLRKQKMLALLETHTVE 286

RESULT 12  
 AAU35619  
 ID AAU35619 standard; Protein, 296 AA.

XX AAU35619;  
 XX 14-FEB-2002 (first entry)  
 XX Haemophilus influenzae cellular proliferation protein #260.  
 XX Antisense; prokaryotic cellular proliferation protein;  
 XX antibiotic; antibacterial; drug design.  
 XX Haemophilus influenzae.  
 XX WO200170955-A2.  
 XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.  
 XX 21-MAR-2000; 2000US-191078P.  
 XX 23-MAY-2000; 2000US-206848P.  
 XX 26-MAY-2000; 2000US-207727P.  
 XX 23-OCT-2000; 2000US-242578P.  
 XX 27-NOV-2000; 2000US-253625P.  
 XX 22-DEC-2000; 2000US-257931P.  
 XX 16-FEB-2001; 2001US-269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Haeselbeck R, Ohlsen KU, Zyekind JW, Wall D, Trawick JD, Carr GJ;  
 XX Yamamoto RT, Xu HH;  
 XX WPI; 2001-611495/70.  
 XX N-PSDB; AAS53478.  
 XX New polynucleotides for the identification and development of  
 XX antibiotics, comprise sequences of antisense nucleic acids -  
 XX Example 3; Seq ID No 11212; 511bp; English.

The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 296 AA;

Query Match 15.0%; Score 368; DB 22; Length 296;  
 Best Local Similarity 36.8%; Pred. No. 4.6e-25;  
 Matches 81; Conservative 45; Mismatches 80; Indels 14; Gaps 3;

QY 25 LDPSFISMNETPOY-DNLQGYAETLERARSKACDESVITGEGTVEGIPVAVILSDFS 83  
 DB 79 LEFDIILKFKOLKXKYKORINAAQKETEK-----DALITMGTLNNPVIIVAAASF 130  
 QY 84 FLGSLGTASVIRIMKAIHRAATELKLPLVSPASGARMOEDNRAFMVMTSITAAVORHR 143  
 DB 131 FMGSSMGSVGAKVKAIAEKRAMENCFVCPFSASGARMOEDNRAFMVMTSITAAVOR 190  
 QY 144 EAHLPFLVLYLNPTMGAMASWSSGHLTPAEPQAQIGFLGPRVVELTTGHALPDGVQA 203  
 DB 191 EKVGFISVLTDPPTGGVTASFSALGDIILAEPSLIGFAGRRVIEQTVREKLPFGQRS 250  
 QY 204 ENLVKTGVIDGIVSPDLQRAAVAKTLKVIQPEATDFSP 243  
 DB 251 EFLLEKGAIDMIVKRSERQTLASVLS-----KLTNQPS 285

RESULT 13  
 ABP28019  
 ID ABP28019 standard; Protein, 291 AA.  
 XX ABP28019;  
 XX



RESULT 15

ABU50977 standard; Protein; 279 AA.

AC ABU50977;

DT 08-MAY-2003 (first entry)

DE Helicobacter pylori selected interacting domain (SID) protein #320.

KW Protein-protein interaction; ulcer; selected interacting domain; SID.

OS Helicobacter pylori.

PN WO200266501-A2.

PD 29-AUG-2002.

PF 28-DEC-2001; 2001WO-EP15428.

PR 02-JAN-2001; 2001US-259302P.

PA (HYBR-) HYBRIGENICS.  
(INSP) INST PASTEUR.

PI Legrain P, Rain J, Colland F, De Reuse H, Labigne A;

DR WPI; 2002-674910/72.

DR N-PSDB; ABX65721.

PT New complexes of protein-protein interactions in Helicobacter pylori,  
PT useful for identifying modulating compounds for treating or preventing  
PT ulcers in mammals -

PS Claim 6; Page 164; 642pp; English.

CC The invention describes a complex of protein-protein interactions in  
CC Helicobacter pylori selected from 421 complexes given in the  
CC specification. The complex of protein-protein interactions are useful  
CC for screening for agents which modulate the interaction of proteins.  
CC Modulating compounds which binds to a targeted bacterial protein may be  
CC used for treating or preventing ulcers in a human or animal. This  
CC is the amino acid sequence of a selected interacting domain (SID),  
CC identified via protein-protein interactions.  
CC Note: Where the patent number printed at the top of the pages in the  
CC specification has obscured areas of protein sequence, the indexer  
CC has replaced the residue with an X to represent an illegible residue.

SQ Sequence 279 AA;

Query Match 14.8%; Score 362; DB 23; Length 279;

Best Local Similarity 39.6%; Pred. No. 1.5e-24;  
Matches 89; Conservative 40; Mismatches 80; Indels 16; Gaps 5;

QY 21 IDSVLDPPDSFISWNE--TPQYDNIN---QGYAETLERARSKAKDESVITGEGTVEGIP 74

DB 56 IEFLLDVGSFEFEFDGRLRPN-DPLNXXVDKESYKQRIKKYEKRTNRPSVISGXAKINRMP 114

QY 75 VAVIISDFSGFLGSGTAVSVIMKAIIRATLKLPLVSPASGARGMOEDNRFAFMMVS 134

DB 115 LQ-IFWDFSPMGSGSVEXEKIVRAINRAVAKREALLVISASGARGMOESTYSIMOMAK 173

QY 135 ITAAVORREAHLPFLVLYRNPTMGAGAMASGSGHLTFAPBPAQIGFLGPRVVELTTGH 194

DB 174 TSAALNRLEAQLPFLISLSDPTTGGVSGSFAFLGDLIAEPGAMIGFAGPRVIVKOTIGA 233

QY 195 ALPDGVQAENLVKTGVIDGIVSPLQLRANA-----KTLKV 231

DB 234 DLPEGFQTAFLHGLIDIMIVRRKDLKTKLSDLIAMMTHTKTSKI 278

Search completed: November 13, 2003, 06:40:21  
Job time : 80 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 06:17:07 / Search time 38 Seconds

(without alignments)  
1242.601 Million cell updates/sec

Title: US-10-024-370-3

Perticle score: 2448

Sequence: 1 VEKREPTVMWGMHTSALT.....NNPERAGRDSRFRERLAQ 491

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1047	42.8	495	2	E70783
2	431.5	17.6	301	2	AE3504
3	392.5	16.0	316	2	AE2101
4	387.5	15.8	289	2	E71875
5	384	15.7	289	2	F64638
6	375	15.3	304	1	AD0802
7	374	15.3	304	1	XMECBD
8	374	15.3	304	2	AE5873
9	374	15.3	304	2	H91028
10	372.5	15.2	284	2	A75422
11	370.5	15.1	282	2	F84045
12	369.5	15.1	276	2	F70340
13	368	15.0	288	2	C86722
14	368	15.0	296	1	B64113
15	363	14.8	262	2	G70001
16	361.5	14.8	298	2	AE2579
17	361.5	14.8	298	2	D97361
18	361.5	14.8	308	2	B71534
19	361.5	14.8	308	2	C86498
20	361.5	14.8	308	2	C72124
21	361	14.7	283	2	H81687
22	361	14.7	411	2	T07261
23	359.5	14.7	326	1	J01238
24	355.5	14.5	280	2	G81429
25	346.5	14.2	290	2	E81171
26	345.5	14.1	290	2	C81934
27	345.5	14.1	304	2	AH0337
28	344	14.1	317	2	AD3003
29	344	14.1	317	2	E98280

30	343.5	14.0	285	2	A89954	acetyl-CoA carboxy
31	336	13.7	288	2	A97338	acetyl-CoA carboxy
32	334	13.6	288	2	D95049	hypothetical prote
33	334	13.6	288	2	B97920	acetyl-CoA carboxy
34	334	13.6	308	2	B82253	acetyl-CoA carboxy
35	332.5	13.6	432	2	T06341	acetyl-CoA carboxy
36	332	13.6	307	2	D87688	hypothetical prote
37	331.5	13.5	489	2	S65564	acetyl CoA carboxy
38	331.5	13.5	507	2	T07012	acetyl-CoA carboxy
39	330.5	13.5	294	2	AG1633	acetyl-CoA carboxy
40	330.5	13.5	515	2	C72344	propionyl-CoA carboxy
41	329.5	13.5	290	2	C83257	acetyl-CoA carboxy
42	329	13.4	288	2	S73119	acetyl-CoA carboxy
43	327	13.4	516	2	A69527	methylnonyl-CoA
44	326.5	13.3	512	2	A05196	hypothetical prote
45	323.5	13.2	493	2	S78384	acetyl-CoA carboxy

## ALIGNMENTS

RESULT 1	
E70783	probable acd3 protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis	
C:Date: 17-Jul-1998	#sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70783	
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S. S.	
i Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentile, S.; Hamilton, N.; Holroyd, S.	
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.	
Nature 393, 537-544, 1998	
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.	
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome	
A:Reference number: AV0500; MUID:96295987; PMID:9634230	
A:Status: preliminary; nucleic acid sequence not shown; translation not shown	
A:Molecule type: DNA	
A:Residues: 1-495 <COL>	
A:Cross-references: GB:273101; GB:AL123456; NID:93261565; PIDN:CAA97379.1; PID:G1314044	
A:Experimental source: strain H37Rv	
C:Genetics:	
A:Gene: acd3	
C:Superfamily: propionyl-CoA carboxylase beta chain	
Query Match	42.8%; Score 1047; DB 2; Length 495;
Best Local Similarity	47.8%; Pred. No. 1.2e-57;
Matches 233; Conservative 73; Mismatches 153; Indels 28; Gaps 8;	
QY	23 SVLPDPSFISWNETPOYDNLNGVAFETLERASRAKODESVITGEGTVPVAVIISDF 82
DB	12 AYLDKGSFVSWDEPLAVPVADSYARELAARAAATGADSVQTBGRVGRRAVAVACEF 71
QY	83 SFLGSLGTVAIVIRKAIHRATEIKLPLVSPASGARMQEDNRAFAVMVSYTAAYQRH 142
DB	72 DFLGSGIGVAAAEIRITAAVERATARPLPLASPSGGRMGEGVAFQWVKIAAIOIOLH 131
QY	143 REAHLPFLVYIRANPTMGAMASGSGHLTPAEQAOIGTGLPRAVEITTHALPDGVQO 202
DB	132 NQARLPVLYVIRHPTTGCVFASWCSLGHITVAEFGALIGFGLPRVYELLYDPPPSGVQT 191
QY	203 AENIVKGVINDIGIYSPLOBRAAVAKTLKVI---QPEVATNRFSPPTPGVA--LPWMEAI 256
DB	192 AENIRRHGIDGVVALDRLPMLRALVTLIDAEPELPA---PQTPAPVPDVPMTDSV 246
QY	257 ARSDPQRFGEIEMETLGADVVLKSGARAGALSFAVRVALARIGRPVVLIGODRRF-- 314
DB	247 VASRRPDRPGVQQLRLRGATDRVLSGDPGEAATTL-LALARRGGQPTVLGGQRAVCG 305
QY	315 ---TLGQELRPARGISLARLNLPIVSIITDSGAEISQAAEELGIASSIARTLSKIID 371
DB	306 GGSITVGPAALREARRGMAAEELCTPLVLYVDAAGPALSAAEQGLAGQIAHCLAEIVT 365
QY	372 AFLPTVSIIGGVGGGALMLPADLVYAENMNLSLAPPEGAGAILPRTINHAETIER 431

Db 366 LDITVTSILGCGSGGRLAALMLPADRVLAALHGMALPPEPGASAIYRDTAAHAAELAA 425  
Qy 432 QGVQAHALLSGGLIDGIVAEFTEHVEEILG---TISNALSELDN-----NPERAGRDSR 482  
Db 426 QGIRASADLLKSGIVDTIVPEYPPDADEBIEFALRLSNIAAEVHALRKIPAPERLA--TR 483  
Qy 483 FTFRPERL 489  
Db 484 LQRYRRI 490

## RESULT 2

AF3504  
acetyl-CoA carboxylase (EC 6.4.1.2) [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis  
C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 03-Jun-2002  
C:Accession: AF3504  
R:DelVecchio, V.G.; Kapetrail, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
Mazur, M.; Golestan, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Levese  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
A:Reference number: AD3252; PMID:11756688  
A:Accession: AF3504  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-301 <KUR>  
A:Cross-references: GB:AE008917; PIDN:AAL53201.1; PID:g17984075; GSPDB:GN00190  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BMEI2020  
A:Map position: 1  
C:Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain  
C:Keywords: ligase

Query Match 17.6%; Score 431.5; DB 2; Length 301;  
Best Local Similarity 43.3%; Pred. No. 1e-19;  
Matches 88; Conservative 43; Mismatches 69; Indels 3; Gaps 2;

Qy 46 YAEFLERARSKAKCDESVITGEGTVEGIPAVILSDPSFLGSLCTVASVRIMKAIHRAT 105  
Db 93 YIDRLKDYRSRGTMDAIVNGITIEGLPIVATVODFSFGMSGLMGAGAILQGFERAI 152  
Qy 106 ELKLPILVSPASGARMQEDNRAPFVMMVSIPTAQRHRAHLPFVYLRLNPMGAMSM 165  
Db 153 ELKRLVLPASSGARMQEGIIISLMQLPRTVAVMLEKAGLPYIVLVLPFTGVTASY 212  
Qy 166 GSSGHLTFAPGAGQIGFLPRVVELTTGHALPDGVOAENLVKTVIDGIVSPQLRAAV 225  
Db 213 AMLGDIHIAEPALIGFAPRIVIEGTIRKLPPEGQSEYLMHEHGMVMVSRLELKATI 272  
Qy 226 AKTKVIT--QPVETDRSPSTTP 246  
Db 273 ARLLKIMTKQPAN-SDAPAPQKP 294

## RESULT 3

AE2101  
acetyl-CoA carboxylase beta chain [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C>Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C:Accession: AE2101  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itiguchi,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium Ana  
baena  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AE2101  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-316 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BAW74063.1; PID:g17131456; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: accD

C:Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain

Query Match 16.0%; Score 392.5; DB 2; Length 316;  
Best Local Similarity 38.0%; Pred. No. 2.8e-17;  
Matches 87; Conservative 42; Mismatches 81; Indels 19; Gaps 3;

Qy 32 SNNETPOYDNL-----NQYAEFLERARSKAKCDESVITGEGTVEGIPAVILSD 81  
Db 85 TWR--PMDENILRATDPLQFRDRKAYSDRLREMEKGLLDVAVTGGQINSSVALAVMD 142  
Qy 82 FSPFLGSLGTVASVRIMKAIHRATELKLPLVSPASGARMQEDNRAPFVMMVSIPTAQR 141  
Db 143 FRFVGSMSGVGEEKIRLIEQATQRRYPVITCTSGARMQEGMSIMQAKISALLER 202  
Qy 142 HREAHLPFLVYLRNPTWGAMASGSGHLTFAPGAGQIGFLPRVVELTTGHALPDGVQ 201  
Db 203 HRDARLLIYIVLNPTTGGVTASFAMGDIILAEPRATIGFAGRRIEQLREKLPDDFQ 262  
Qy 202 QAENLVKTGVIDGIVSPQLRAAVAKTKIYQVEATDRSPSTTPGVAL 250  
Db 263 TAEDLLKGFVDDIVPRTQLKNTLSQILAHQRY-----PTTPPVVL 304

## RESULT 4

E71875  
acetyl-coenzyme A carboxylase chain B - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 16-Jul-1999

C:Accession: E71875  
R:Alm, R.A.; Gibson, L.S.L.; Moit, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
Ives, C.; Gibson, R.; Metters, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; J  
Nature 397, 176-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric patho  
gen Helicobacter pylori  
A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: E71875  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-289 <ABN>  
A:Cross-references: GB:AE001518; GB:AE001439; NID:g4155454; PIDN:AAD06468.1; PID:g4155465

A:Experimental source: strain J99  
C:Genetics:  
A:Gene: accD  
C:Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain

Query Match 15.8%; Score 387.5; DB 2; Length 289;  
Best Local Similarity 39.6%; Pred. No. 5.2e-17;  
Matches 91; Conservative 43; Mismatches 81; Indels 15; Gaps 4;

Qy 16 SALTLIDSVLDPSFISMN--TFQYDNLN---QGYAEFLERARSKAKCDESVITGEGT 69  
Db 60 NATERIEFLCDVGSFEEFDKLRPN-DPLNFVDSEYKORIKKYEKRTNPSSVISEAK 118  
Qy 70 VEGIPAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGARMQEDNRAPF 129  
Db 119 INRNPGLIVAFDFFSFGMSLGSVEGKIVRINRAVAKKELLVASGARMQESTYSL 178  
Qy 130 VMMVSIPTAQRHRAHLPFVYLRLNPTWGAMASGSGHLTFAPGAGQIGFLPRVVE 189  
Db 179 MQMKTSAALNRLEALPFLSLSDPTTYGGVSAFPLGDLIIAEFGAMIGFAPRIVIK 238  
Qy 190 LTTGHALPDGVOAENLVKTVIDGIVSPQLRAAV-----KTLKV 231  
Db 239 QTIGADLPFGQTAEFLIEHGLIDMIVHRKDLKKTSLDLIAMMTHKTSKI 288

## RESULT 5

F64638  
acetyl-CoA carboxylase beta subunit - Helicobacter pylori (strain 26695)  
C:Species: Helicobacter pylori







Qy 130 VMWSTTAHVORRHRAHLPFLVYLNRPTNGAMASWGSGHLPFAEPGAOIGFLGRVVE 189  
 Db 175 MOKAKTVALTALSGRGVYLSLTPDTGGVATADVIARFGLIGAGRGVAC 234  
 Qy 190 LTTGHALPDGVQOAEHLVKTGVISPLQRAAVALKLYIQPEA 237  
 Db 235 QTRONLPBGFQRAEFLLEHGMVDVNRREORRHLAGLGLVLTROEA 282

## RESULT 11

F84045  
 acetyl-CoA carboxylase transferase beta subunit accd [imported] - *Bacillus halodurans* (S  
 C)Species: *Bacillus halodurans*  
 C)Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C)Accession: F84045  
 R)Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Maeni, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A)Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
 A)Reference number: A83650; MUID:20512582; PMID:11058132  
 A)Accession: F84045  
 A)Status: preliminary  
 A)Molecule type: DNA  
 A)Residues: 1-282 <STO>  
 A)Cross-references: GB:AP001517; GB:BA000004; NID:g10175500; PIDN:BA06885.1; GSPDB:GN00  
 A)Experimental source: strain C-125  
 C)Genetics:  
 A)Gene: accd  
 C)Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain

Query Match 15.1%; Score 370.5; DB 2; Length 282;  
 Best Local Similarity 36.5%; Pred. No. 5.7e-16;  
 Matches 85; Conservative 42; Mismatches 85; Indels 21; Gaps 3;

Qy 11 GMEH-TSALTLDVLPDPSFISMETPQYDLNCG-----VATETERRASRK 58  
 Db 55 GFHHRMNAFDRLESLLDPGTFV-----ELDKMTTEDEPLSPFYREKVEADRKKTN 105  
 Qy 59 CDESVITGEVTEGIPVAVILSDFSFLGSLGTVAASVIRIMKAIHRATELKLPLVSPASG 118  
 Db 106 LNEALVTGEETNGFPVLVGVNDARFRKSGMSVSGEKTRAIERAIENKQFPLPFSAG 165  
 Qy 119 GARMOEDNRFAVMVSTAAVORRHRAHLPFLVYLNRPTNGAMASWGSGHLPFAEPGA 178  
 Db 166 GARMOEGVLSLQMAKTSALLERLDRVGLFISVMTHTPTGGVSASFSLGYNFAEPGA 225  
 Qy 179 QIGFLGRVVELTTGHALPDGVQOAEHLVKTGVISPLQRAAVALKLY 231  
 Db 226 LIGFAGRRITIEQTIREELPNDPQTAEFLKQQLDRVIPRSEMKDTLTILDI 278

## RESULT 12

F70340  
 acetyl-CoA carboxyltransferase beta subunit - *Aquifex aeolicus*  
 C)Species: *Aquifex aeolicus*  
 C)Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 16-Jul-1999  
 C)Accession: F70340  
 R)Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov  
 V.  
 Nature 392, 353-358, 1998  
 A)Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.  
 A)Reference number: A70300; MUID:98196666; PMID:9537320  
 A)Accession: F70340  
 A)Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A)Molecule type: DNA  
 A)Residues: 1-216 <AQR>  
 A)Cross-references: GB:AE000690; NID:g2983100; PIDN:AA006712.1; PID:g2983107; GB:AE00065  
 A)Experimental source: strain VFS  
 C)Genetics:  
 A)Gene: accd  
 C)Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain  
 Query Match 15.1%; Score 369.5; DB 2; Length 276;  
 Best Local Similarity 38.7%; Pred. No. 6.4e-16;

Matches 86; Conservative 42; Mismatches 89; Indels 5; Gaps 2;  
 Qy 16 SALTLIDSLDPS-FISMNTERPOYDNN-----OGVATLEBARAKACDESVINGEGV 70  
 Db 50 SALBKIKYTLDBENPELLFEETLPADPLNPKDTKSKYKORIKKAOGETGISEAIVTEGHI 109  
 Qy 71 EGIPIAVILSDPSFLGSLGTVAASVIRIMKAIHRATELKLPLVSPASGARMOEDNRFAV 130  
 Db 110 KGRALIAVMDFNFIQSGMSGVGERFPACRAVAETHTPLISFASGARMOEGISLM 169  
 Qy 131 VMWSTTAHVORRHRAHLPFLVYLNRPTNGAMASWGSGHLPFAEPGAOIGFLGRVVE 190  
 Db 170 QMAKTFGVGLLKEANIPYISVLTNPITGVASAPFGLDILABPKALIGPAGRVIEQ 229  
 Qy 191 TTHGHALPDGVQOAEHLVKTGVISPLQRAAVALKLY 232  
 Db 230 TTKQKLPEGFQYAEFLLEKQIDMVHRRKELDKLAFLEVM 271

## RESULT 13

C86722  
 hypothetical protein accd [imported] - *Lactococcus lactis* subsp. *lactis* (strain IL1403)  
 C)Species: *Lactococcus lactis* subsp. *lactis*  
 C)Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C)Accession: C86722  
 R)Bolotin, A.; Wincker, P.; Mauer, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrli  
 Genome Res. 11, 731-753, 2001  
 A)Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ss  
 A)Reference number: A86625; MUID:21235186; PMID:11337471  
 A)Accession: C86722  
 A)Status: preliminary  
 A)Molecule type: DNA  
 A)Residues: 1-288 <STO>  
 A)Cross-references: GB:AE005176; PID:g12723698; PIDN:AAK04877.1; GSPDB:GN00146  
 A)Experimental source: strain IL1403  
 C)Genetics:  
 A)Gene: accd  
 C)Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain

Query Match 15.0%; Score 368; DB 2; Length 288;  
 Best Local Similarity 40.0%; Pred. No. 8.4e-16;  
 Matches 86; Conservative 37; Mismatches 90; Indels 2; Gaps 2;

Qy 24 VLDPDSFISMETPQYD-LN-OGVATLEBARAKACDESVINGEGVPAVILSD 81  
 Db 72 VADKDSFVEMFTGIESKNPLDPGIPKTLAAATKATGIDEAVITGTATIKCOKTALATMD 131  
 Qy 82 FSLGSLGTVAASVIRIMKAIHRATELKLPLVSPASGARMOEDNRFAVMVSTAAVOR 141  
 Db 132 STFIASMGIVGELTRLPLFYATTEKLPITVTRASGARMOEGISLMQMAKTSAAVYR 191  
 Qy 142 HREAHLPFLVYLNRPTNGAMASWGSGHLPFAEPGAOIGFLGRVVELTTGHALPDGVQ 201  
 Db 192 HSNAGLFYITVLTPTTGGVTSAPSLDIDILAEPOSILGAGRVIRQYRQTLPPDFO 251  
 Qy 202 QAEHLVKTGVISPLQRAAVALKLYIQPE 236  
 Db 252 KAEPFLNGFVDAIVKTELKQKLALLLEHTVE 286

## RESULT 14

B64113  
 acetyl-CoA carboxylase (EC 6.4.1.2) carboxyltransferase beta chain - *Haemophilus influen*  
 C)Species: *Haemophilus influenzae*  
 C)Date: 18-Aug-1995 #sequence\_revision 23-Feb-1996 #text\_change 18-Jun-1999  
 C)Accession: B64113  
 R)Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A  
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J  
 ; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fritchman, J.L.; Geoghegan, N.S.M.  
 Science 269, 496-512, 1995  
 A)Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
 A)Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.  
 A)Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: B64113  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-236 <TIGR>  
A:Cross-references: GB:U32806; GB:L42023; NID:g1574717; PIDN:AAC22913.1; PID:g1574718; T  
A:Note: named as homolog to a protein from *Escherichia coli*  
C:Function:  
A:Description: the enzyme complex catalyzes the synthesis of malonyl-CoA; the overall re  
action bound to BCCP (see BKE9) with bicarbonate; carboxyltransferase catalyzes the trans  
A:Pathway: fatty acid biosynthesis  
A:Note: first committed step  
C:Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain  
C:Keywords: fatty acid biosynthesis; ligase; zinc finger  
F:29-31/Region: zinc finger CCCC motif

Query Match 15.0%; Score 368; DB 1; Length 296;  
Best Local Similarity 36.8%; Pred. No. 8.7e-16; Indels 14; Gaps 3;  
Matches 81; Conservative 45; Mismatches 80;

QY 25 LDDPSFISWNETPOY-DNLNQGVAETLERARSKACDESVITGEGVEGIPVAVILSDPS 83  
DB 79 LEPRDLIKFDLKKYDRIIAAOKETGEK-----DALITMTGTLNMPVIVVAAASNFA 130  
QY 84 FLGSLIGTVA SVIRIMKAIHRATELKLPLVSPASGARMQEDNRAFMVMSITRAVQRHR 143  
DB 131 FMGSGMSGVGAKFVKAEKAMENWCPFCFSASGARMQEDALFSLQMAKTSAVLAQMR 190  
QY 144 EAHLPPLVYIRNPTMGMAWSGSSGHLTPAEPGAQIGFLGPRVVELTGTALPDGVQQA 203  
DB 191 EKGVPFISVLTDPITGVSASFAMLDGDLNIAEPKALIGFAGPRVIEQTVREKLEPEGQRS 250  
QY 204 ENLVKTVGIDIVSPLOLRAAVAKTLKVIOPEATDRFSP 243  
DB 251 EFLLEKGAIDMIVRSEMRQTASVLS-----KLITNPSP 285

## RESULT 15

G70001  
acetyl-CoA carboxylase homolog yctI - *Bacillus subtilis*

C:Species: *Bacillus subtilis*  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000

C:Accession: G70001  
R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle  
lech, U.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
Y., M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
A:Authors: Schleicher, S.; Schoeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumestein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: G70001  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-262 <KUN>  
A:Cross-references: GB:Z99119; GB:AI009126; NID:g2635200; PIDN:CAB14881.1; PID:g2635386  
A:Experimental source: strain 168  
C:Genetics:

A:Gene: yctI  
C:Superfamily: acetyl-CoA carboxylase, carboxyltransferase beta chain

Query Match 14.8%; Score 363; DB 2; Length 262;  
Best Local Similarity 38.2%; Pred. No. 1.5e-15;  
Matches 84; Conservative 41; Mismatches 89; Indels 6; Gaps 2;

QY 21 IDSVLDDPSFISWNETPOYDNL--NQGVAETLERARSKACDESVITGEGVEGIPVAVI 78

DB 38 IESIMDEGSFEEFNOQMSENPLGEPGLLEKLEKDRKTSINAEVAVCKGIGGHPAVVA 97  
QY 79 LSDPSFLGSLIGTVA SVIRIMKAIHRATELKLPLVSPASGARMQEDNRAFMVMSITRA 138  
DB 98 VMDSSFRMGSGVSGEKTITLIERAKADKVPFIIFTSAGARMQEGVLSIMQAKTSSA 157  
QY 139 VORHREAHLPPLVYIRNPTMGMAWSGSSGHLTPAEPGAQIGFLGPRVVELTGTALPD 198  
DB 158 IKLFSFEEQGLIISVMTHTPTGVSASFASLDGYNFAEPGALIGFAGPRVIEQTVREKLE 217  
QY 199 GVQQAENLVKTVGIDIVSPLOLRAAVAKTLKVIOPEAT 238  
DB 218 DQTAEPFLIKHQDLDAVH----RDMKKTLENLMDHQT 253

Search completed: November 13, 2003, 06:45:04  
Job time : 39 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 02:26:42 / Search time 24 Seconds

(without alignments)  
962.088 Million cell updates/sec

Title: US-10-024-370-3

Perfect score: 2448

Sequence: 1 VEXRPPTVMGMEHTSALT.....NNPERAGRDSRFRPERLAQ 491

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1047	42.8	495	1	ACCD MYCTU
2	389	15.9	305	1	ACCD SYN7
3	374	15.3	304	1	ACCD ECOLI
4	368	15.0	296	1	ACCD HAEIN
5	361	14.7	411	1	ACCD CHLUV
6	359.5	14.7	326	1	ACCD SYN3
7	352.5	13.6	432	1	ACCD SOYBN
8	331.5	13.5	489	1	ACCD BRANA
9	330.5	13.5	501	1	ACCD LOTVA
10	329	13.4	288	1	ACCD PORPU
11	326.5	13.3	512	1	ACCD TOBAC
12	324.5	13.3	488	1	ACCD ARATH
13	323.5	13.2	493	1	ACCD EPIV1
14	322	13.2	314	1	ACCD ANGLY
15	321.5	13.1	522	1	ACCD SPIOL
16	318.5	13.0	321	1	ACCD PINTH
17	316.5	12.9	590	1	ACCD TEA
18	315	12.9	316	1	ACCD MARPO
19	313.5	12.8	497	1	ACCD CUSRE
20	311.5	12.7	438	1	ACCD OENHO
21	310	12.7	313	1	ACCD ANFTO
22	302	12.3	506	1	PCCB BACSV
23	300	12.3	319	1	ACCD PICAB
24	288	11.8	315	1	ACCA HAEIN
25	284	11.6	315	1	ACCD PHYP
26	283.5	11.6	546	1	PCCB SACR
27	278.5	11.4	324	1	ACCA CYACA
28	269	11.0	318	1	ACCA SALTY
29	266	10.9	318	1	ACCA ECORI
30	264	10.8	539	1	PCCB HUMAN
31	261.5	10.7	323	1	ACCA AQUAN
32	261	10.7	539	1	PCCB PIG
33	260.5	10.6	325	1	ACCA_BACSV

34	259.5	10.6	326	1	ACCA SYN3
35	257	10.5	312	1	ACCA HELPY
36	257	10.5	312	1	ACCA HELPY
37	254	10.4	541	1	PCCB RAT
38	252.5	10.3	327	1	ACCA SYN7
39	251	10.3	548	1	PCCS MYCTU
40	248.5	10.2	322	1	ACCA ANTSP
41	246.5	10.1	549	1	PCCB MYCBE
42	238	9.7	324	1	ACCA PORPU
43	215.5	8.8	476	1	PCCB THOER
44	204.5	8.4	473	1	PCCB MYCTU
45	190.5	7.8	563	1	MCCB_HUMAN

## ALIGNMENTS

RESULT 1  
ACCD MYCTU  
ID ACCD MYCTU STANDARD; PRT; 495 AA.  
AC 010561:  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Putative acetyl-coenzyme A carboxylase subunit  
DE beta (BC 6.4.1.2) (ACCase beta chain).  
GN ACCD3 OR RV0904C OR MT0927 OR MTCY31.32C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Kettlewell J., Mclennan J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulten J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544 (1998).  
[12]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fletschman R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A.L., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains." to the EMBL/GenBank/DBJ databases.  
RL Submitted (Apr-2001)  
CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A  
CC CARBOXYLASE COMPLEX; FIRST, BIOTIN CARBOXYLASE CATALYZES THE  
CC CARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE  
CC TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: ADP + acetyl-CoA + HCO(3) (-) = ADP + phosphate  
CC + malonyl-CoA.  
CC -!- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: 273101; CAA97379.1; -  
 DR EMBL: AEO06979; AAK45174.1; -  
 DR PIR: E70783; E70783.  
 DR TIGR: MT0927; -  
 DR Tuberculae; Rv0904c; -  
 DR Interpro; IPR000438; ACACCC\_transf.  
 DR Interpro; IPR000022; Carboxyl\_trans.  
 DR Pfam; PF01039; Carboxyl\_trans; 1.  
 DR PRINTS; PR01070; ACCCTFRASB.  
 DR Hypothetical protein; Fatty acid biosynthesis; ligase;  
 KM Complete proteome.  
 SQ SEQUENCE 495 AA; 51772 MW; 156C96B27FF950C CRC64;  
 Query Match 42.8%; Score 1047; DB 1; Length 495;  
 Best local Similarity 47.8%; Pred. No. 1,le-54;  
 Matches 233; Conservative 73; Mismatches 153; Indels 28; Gaps 8;  
 QY 23 SVLDPPDSFISWNETPOYDNLNOCYAEVLEERAKSKACDESVITGEGIVEGIPAVILSDF 82  
 DB 12 AVLDKGSFVSMDEPLAVPADSYARELAARAATGADESVGTGEGVFRRAVAVACEF 71  
 QY 83 SFLGSGIGTAVSVIRKAIHRAATELKLPLVSPASGARMOEDNRAFMVMSITAAVORH 142  
 DB 72 DFLGSGIGVAAAEKITAVERATKERPLPLASPGSGTRQEGVAFLOVVKIAAATQLH 131  
 QY 143 REAHLPELVYLRNPTMGAMASWSSGHLTFAEFGAIGFLGPRVVELTGHALPDGVQO 202  
 DB 132 NQRLPLVLYLRHPTTGGVFAWSGSLGHLTVAEFGALIGFLGPRVVELTGDPPSGVQO 191  
 QY 203 AENLVKTVGVDIGVSPLOLRAAVAKTLKVI---QPEARDRSPPTPGVA--LPVMEAI 256  
 DB 192 AENLVKTVGVDIGVSPLOLRAAVAKTLKVI---QPEARDRSPPTPGVA--LPVMEAI 256  
 QY 257 ARSRDPORPGIGEMETLGAADVVLKSGARAGALSPAVVALARIGAPVVLIGDRPF-- 314  
 DB 247 VASRPDRPGVROLRRHGAIDRVLLSGTDEGAATVL-LALARFGGPTVVLGGQRAVGG 305  
 QY 315 ---TLGQELRPARRGISLARLNLPIVSTIDTSGAEISQAAEELGIASSARTLSKID 371  
 DB 306 GGSITVGPAALREARRGMAALAECLPLVLVIDAAGPLSAAECCGLAGLHCLAEVLVT 365  
 QY 372 APLPTVVIIGGGVGGGGLAMLPADLVYAENAMVLSALPPEGASAILPRTNHAETIER 431  
 DB 366 LDTPTVSLILGSGGGPALAMLPADRVLAALHGLALPPEPGASAILVTRDRAHAELAA 425  
 QY 432 QGVAAHLLSOGLLDGIVAETEHFVEELIG---TISNALSELD-----NPERAGDSR 482  
 DB 426 QGIRSDLLKSGIVDTIVPEYPADEDEIEFALRLSNIAAEVAHAKRIPAPERLA--TR 483  
 QY 483 FTRPERL 489  
 DB 484 LQRYRRI 490  
 RESULT 2  
 ACDD\_SYNPF STANDARD; PRT; 305 AA.  
 ID ACDD\_SYNPF  
 AC 054776;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta  
 DE (EC 6.4.1.2) (ACCase beta chain).  
 GN ACDD.  
 OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.  
 OX NCBI\_TaxID=1140;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Phung L.T., Haselkorn R.;  
 RL Submitted (JUL-1996) to the EMBL/Genbank/DDBJ databases.  
 CC FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A  
 CC CARBOXYLASE COMPLEX; FIRST, BIOTIN CARBOXYLASE CATALYZES THE

CC CARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE  
 CC TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate  
 CC + malonyl-CoA.  
 CC -1- SUBUNIT: ACETYL-COA CARBOXYLASE IS AN HETEROHEXAMER OF BIOTIN  
 CC CARBOXYL CARRIER PROTEIN, BIOTIN CARBOXYLASE AND THE TWO SUBUNITS  
 CC OF CARBOXYL TRANSFERASE IN A 2:2 COMPLEX (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE ACDD / PCOB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U59237; AAB82034.1; -  
 DR Interpro; IPR000438; ACACCC\_transf.  
 DR Interpro; IPR000022; Carboxyl\_trans.  
 DR Pfam; PF01039; Carboxyl\_trans; 1.  
 DR PRINTS; PR01070; ACCCTFRASB.  
 DR TIGR; TIGR00515; accd; 1.  
 KM Fatty acid biosynthesis; ligase; zinc-finger.  
 FT ZN FING 33  
 SQ SEQUENCE 305 AA; 33807 MW; 9AFCF75A8E9F3B5A CRC64;  
 Query Match 15.9%; Score 389; DB 1; Length 305;  
 Best local Similarity 35.5%; Pred. No. 3,le-16;  
 Matches 86; Conservative 50; Mismatches 90; Indels 16; Gaps 2;  
 QY 21 IDSVLPDSFISWNETPOYDNLNOCYAEVLEERAKSKACDESVITGEGIVEGIPAVILS 80  
 DB 79 LDEAVSFTDPLGRD-----RKSYSDLRKETQANTGSDAVRTGVLLEGQPAVLGV 131  
 QY 81 DFLGSGIGTAVSVIRKAIHRAATELKLPLVSPASGARMOEDNRAFMVMSITAAVQ 140  
 DB 132 DFRMGSGMSVVGKELRLIEKTEORSPIVIVCASGARMOEGLSLQMKISGALR 191  
 QY 141 RHEAHLPELVYLRNPTMGAMASWSSGHLTFAEFGAIGFLGPRVVELTGHALPDGV 200  
 DB 192 RHEAHLPELVYLRNPTMGAMASWSSGHLTFAEFGAIGFLGPRVVELTGHALPDGV 200  
 QY 201 QQAEVLKTVGVDIGVSPLOLRAAVAKTLKVIOPEVATDRFSPPTPGVALPVEAIARSR 260  
 DB 252 QTAEVLQAHGFVDTIVPRQLKKTIAQLRLHQ-----PQSPKMLPLESSSPAR 302  
 QY 261 DP 262  
 DB 303 AP 304  
 RESULT 3  
 ACDD\_ECOLI STANDARD; PRT; 304 AA.  
 ID ACDD\_ECOLI  
 AC P08193; P76937; P78251;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta  
 DE (EC 6.4.1.2) (ACCase beta chain).  
 GN ACDD OR DEDB OR USG OR B2316 OR Z3578 OR ECS3200.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562; 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87308246; PubMed=3040739;  
 RA Bognar A.L., Osborne C., Shane B.;  
 RT "Primary structure of the Escherichia coli folC gene and its  
 RT folY1polY5lutamate synthetase-dihydrofolate synthetase product and

RT regulation of expression by an upstream gene.";  
 RL J. Biol. Chem. 262:12337-12343(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97308226; PubMed=3040734;  
 RA Noret M.L., Marvel C.C., Tolani D.R.;  
 RT "The hst-purF region of the Escherichia coli K-12 chromosome."  
 RT Identification of additional genes of the hst and purF operons.";  
 RL J. Biol. Chem. 262:12209-12217(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97349980; PubMed=9205837;  
 RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitushashi N.,  
 RA Mitobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,  
 RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivaundaram S.,  
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
 RA Yamagata S., Horikuchi T.;  
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
 RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and  
 RT analysis of its sequence features.";  
 RL DNA Res. 4:91-113(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller J.,  
 RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ichii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasekawa C., Ogaawara N., Yasunaga T.,  
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 RN [7]  
 RP SEQUENCE OF 1-31 FROM N.A.  
 RX MEDLINE=93123150; PubMed=7678242;  
 RA Li S.-J., Cronan J.E. Jr.;  
 RT "Growth rate regulation of Escherichia coli acetyl coenzyme A  
 RT carboxylase, which catalyzes the first committed step of lipid  
 RT biosynthesis.";  
 RL J. Bacteriol. 175:332-340(1993).  
 RN [8]  
 RP SIMILARITY TO ZPPA.  
 RX MEDLINE=91360091; PubMed=1886618;  
 RA Nagano Y., Matsuno R., Sasaki Y.;  
 RT "An essential gene of Escherichia coli that has sequence similarity  
 RT to a chloroplast gene of unknown function.";  
 RL Mol. Gen. Genet. 228:62-64(1991).  
 RN [9]

RP FUNCTION.  
 RX MEDLINE=92380957; PubMed=1355086;  
 RA Li S.-J., Rock C.O., Cronan J.E. Jr.;  
 RT "The dedB (usg) open reading frame of Escherichia coli encodes a  
 RT subunit of acetyl-coenzyme A carboxylase.";  
 RL J. Bacteriol. 174:5755-5757(1992).  
 CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A  
 CC CARBOXYLASE COMPLEX. FIRST, BIOTIN CARBOXYLASE CATALYZES THE  
 CC CARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE  
 CC TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA.  
 CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3)(-) = ADP + phosphate  
 CC + malonyl-CoA.  
 CC -1- SUBUNIT: ACETYL-COA CARBOXYLASE IS AN HETEROHXAMER OF BIOTIN  
 CC CARBOXYL CARRIER PROTEIN, BIOTIN CARBOXYLASE AND THE TWO SUBUNITS  
 CC OF CARBOXYL TRANSFERASE IN A 2:2 COMPLEX.  
 CC -1- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by, and for commercial/  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M32445; AAA23807.1; -;  
 DR EMBL: J02808; AAA23801.1; -;  
 DR EMBL: M68934; AAA23965.1; -;  
 DR EMBL: AE000320; AAC75376.1; -;  
 DR EMBL: D90862; BAA16165.1; -;  
 DR EMBL: D90863; BAA16173.1; -;  
 DR EMBL: AE005463; AAG57445.1; -;  
 DR EMBL: AP002561; BAB36623.1; -;  
 DR EMBL: S53037; AAB24894.2; -;  
 DR PIR: A85873; A85873.  
 DR PIR: B65004; XMECBD.  
 DR PIR: H91028; H91028.  
 DR Ecogene: EG10217; accd.  
 DR InterPro: IPR000438; ACOAC.transfb.  
 DR InterPro: IPR000022; Carboxyl\_trans.  
 DR Pfam: PF01039; Carboxyl\_trans.  
 DR PRINTS: PRO1070; ACCCTFRASEB.  
 DR TIGRFAMs: TIGR00515; accd; 1.  
 KW Fatty acid biosynthesis; Ligase; Zinc-finger; Complete proteome.  
 FT ZN\_FING 27 49  
 FT CONFLICT 76 77 EL -> SV (IN REF. 2).  
 FT CONFLICT 225 239 IGFAPRVIEQTVRE -> MALPVRLVSNRPFAK (IN  
 FT REF. 1).  
 FT CONFLICT 226 235 GFAGPRVIEQ -> ALPVRVLSNR (IN REF. 2).  
 FT SEQUENCE 304 AA; 3332 MW; 401FEC94D728F3CB CMC64;  
 QY Query Match 15.3%; Score 374; DB 1; Length 304;  
 QY Best Local Similarity 36.0%; Pred. No. 2.3e-15;  
 QY Matches 85; Conservative 49; Mismatches 92; Indels 10; Gaps 3;  
 DB 17 ALTLIDSLDPSRISNNETPOY-DNLNGVAELTERASAKKDDSEVITSEGVGIPV 75  
 DB 69 SLVELGSELEPKDVLKFRDSKKYDRLASAKETGEK-----DALVWKGTLTYGMPV 120  
 QY 76 AVILDSFSLGSGIGTVASVIRKAIHRATELKLPLVSPASGARQOEDRAFAVMVSI 135  
 DB 121 VAAAFEPFAGSGSVGAFVAVAEQALDENCPICFSASGARQOELMSLMQNAKT 180  
 QY 136 TAAVQRRREAHLPVLVIRNPTGGMASWSSGHLTPAEFGAIGFLGPRVVELTGA 195  
 DB 181 SAALAKQERGLPYISVLTPTMGVSASFAMGLDNLNIAEPKALIGFGPRVIEQTVREK 240  
 QY 196 LPDGVQAAENLVKGVIGDGVSPLOLRAAVAKTLKVIQPVATDRFSPPTGVGLP 251  
 DB 241 LPDFQSEFLIEKGAIDMIVRPEMRUKLASILAKMLPAVPZAP-REGVVVP 295  
 RESULT 4

ACCD\_HAEIN STANDARD; PRT; 296 AA.  
ID ACCD\_HAEIN STANDARD; PRT; 296 AA.  
AC P43779;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta  
DE (EC 6.4.1.2) (ACCase beta chain).  
GN ACCD OR H11260.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McQuay K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd.";  
RL Science 269:496-512(1995).  
CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A  
CARBOXYLASE COMPLEX. FIRST, BIOTIN CARBOXYLASE CATALYZES THE  
CARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE  
TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate  
+ malonyl-CoA.  
CC -1- SUBUNIT: ACETYL-COA CARBOXYLASE IS AN HETEROHEXAMER OF BIOTIN  
CARBOXYL CARRIER PROTEIN, BIOTIN CARBOXYLASE AND THE TWO SUBUNITS  
OF CARBOXYL TRANSFERASE IN A 2:2 COMPLEX (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE ACCD / PCOB FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U32806; AAC22913.1; -.  
DR PIR; B64113; B64113.  
DR TIGR; H11260; -.  
DR InterPro; IPR000438; ACOACC\_transfB.  
DR InterPro; IPR000022; Carboxyl\_trans.  
DR Pfam; PF01039; Carboxyl\_trans. 1.  
DR PRINTS; PRO1070; ACCCTFRASEB.  
DR TIGRFAMS; TIGR00515; accd; 1.  
KW Fatty acid biosynthesis; Ligase; Zinc-finger; Complete proteome.  
FT ZN FING 29 51 C4-TYPE (POTENTIAL).  
SQ SEQUENCE 296 AA; 32635 MW; 3FEE316557FD4553 CRC64;  
Query Match 15.0%; Score 368; DB 1; Length 296;  
Best Local Similarity 36.8%; Pred. No. 5.1e-15;  
Matches 81; Conservative 45; Mismatches 80; Indels 14; Gaps 3;  
QY 25 LDPDSFISWNETPOY-DNINOGYAETLEBARSKACDBSVITGEGTGVIPAVILSPS 83  
DB 79 LEPKOILFKFDLKKYKDRINNAQKETGEK-----DALITMTGLVMPPIVVAASNFA 130  
QY 84 FLGSLGVAVSRIMKAIHRATELKLPLVSPASGAGMOENRNFVMMVSTTAVOHR 143  
DB 131 FMGSGVSGVAKFVYAKKAKEMNNCPVFCFSASGARQDELFLMDQAKTSAVLAQMR 190  
QY 144 EAHLPFLVLYLRNPTMGAMASWSSGHLTFAEPGAQIGFLGPRVVELTGHALPDGVOA 203

DB 191 EKVGFISVLTDPILGVGSASFAMLGDLNIAEPKALIGFAGPRVIEQTVAREKLEPEGFORS 250  
QY 204 ENLVKTVINDIGVSPDLQRAAVAKTLKVIOPVATDRFSP 243  
DB 251 EFLLEKGAIDMIVKRSERQTLASVLS-----KLTVQPS 285  
RESULT 5  
ACCD\_CHLVU STANDARD; PRT; 411 AA.  
ID ACCD\_CHLVU STANDARD; PRT; 411 AA.  
AC P56293;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta  
DE (EC 6.4.1.2) (ACCase beta chain).  
GN ACCD.  
OS Chlorella vulgaris.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;  
OC Chlorellaceae; Chlorella.  
OX NCBI\_TaxID=3077;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IAM C-27 / Tamiya;  
RX MEDLINE=97303241; PubMed=9159184;  
RA Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito M., Ito S.,  
RA Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,  
RA Inamura A., Yoshinaga K., Sugitara M.;  
RT "Complete nucleotide sequence of the chloroplast genome from the  
RT green alga Chlorella vulgaris: the existence of genes possibly  
RT involved in chloroplast division.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).  
CC -1- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE  
CC INVOLVED IN CHLOROPLAST CARBON METABOLISM.  
CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate  
+ malonyl-CoA.  
CC -1- SIMILARITY: BELONGS TO THE ACCD / PCOB FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AB001684; BA57908.1; -.  
DR PIR; T07261; T07261.  
DR InterPro; IPR000438; ACOACC\_transfB.  
DR InterPro; IPR000022; Carboxyl\_trans.  
DR Pfam; PF01039; Carboxyl\_trans. 1.  
DR PRINTS; PRO1070; ACCCTFRASEB.  
DR TIGRFAMS; TIGR00515; accd; 1.  
KW Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.  
FT ZN FING 36 58 C4-TYPE (POTENTIAL).  
SQ SEQUENCE 411 AA; 46547 MW; 81EB041499087E79 CRC64;  
Query Match 14.7%; Score 361; DB 1; Length 411;  
Best Local Similarity 31.7%; Pred. No. 1.9e-14;  
Matches 102; Conservative 60; Mismatches 114; Indels 46; Gaps 7;  
QY 16 SATLISVLDPPDSFISWNE-----TQYDNINOGYAETLEBARSKACDBSVITGEGTV 70  
DB 65 SSTERISLVANDANTRPFDVSPCDLFRDQAYTERLKDAGERTGLDVAVTGGL 124  
QY 71 EGIIVAVILSPFSLGSLGVASVRIMKAIHRATELKLPLVSPASGAGARMQENRAV 130  
DB 125 DGIPIALGVMDFFHMGSGVSGEKITRLILVYTGGLVIVLVCASGARMQGIISLM 184  
QY 131 MMSVITAAVORHRE-AHLPLVLYLRNPTMGAMASWSSGHLTFAEPGAQIGFLGPRVVE 189



Db 185 QWAKISALHONCAKLYISVLTSPPTGVTAFAMLDLLFAEPKALIGFAGRRVIE 244  
 QY 190 LITGHALPDGVOQAENLVKTGVIDGIVSFLQRAVAKTKYIQ--PVEATPRF-----S 242  
 Db 245 QTLQEQLPDDPTATYALHGHGLDIPRSPFLKQALSETLTYLKAPLKEQGRIPYGERG 304  
 QY 243 PTTGVALPVMFAIRSDPQRPGI-GEIMETGADVLTSGARAGALSPAVVALARIQ 301  
 Db 305 PLTKRREGQLRFRLLSKSTPEVILVNDKELG----- 338  
 QY 302 GRPVVLIGQDRRFTLGPQELRF 323  
 Db 339 -----FLGQTFQ-TLYPEKLEF 354

## RESULT 6

ACCD\_SYNY3 STANDARD; PRT; 326 AA.  
 ID ACCD\_SYNY3  
 AC 057417;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DB Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta  
 DE (EC 6.4.1.2) (ACCase beta chain).  
 GN ACCD OR ZPPA OR SLO336.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 CX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=92118327; PubMed=1368738;  
 RA Ogura Y., Yoshida T., Nakamura Y., Takemura M., Oda K., Ohnaka K.;  
 RT "Gene encoding a putative zinc finger protein in Synechocystis  
 PCC6803."  
 RL PCC6803.  
 RL Agric. Biol. Chem. 55:2259-2264 (1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96127529; PubMed=8590279;  
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,  
 RA Sugita T., Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb  
 region from map positions 64k to 92k of the genome."  
 RL DNA Res. 2:153-166 (1995).  
 CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A  
 CARBOXYLASE COMPLEX. FIRST, BIOTIN CARBOXYLASE CATALYZES THE  
 TRANSFER OF THE BIOTIN CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE  
 TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate  
 + malonyl-CoA.  
 CC -1- SUBUNIT: ACETYL-COA CARBOXYLASE IS AN HETEROHEXAMER OF BIOTIN  
 CARBOXYL CARRIER PROTEIN, BIOTIN CARBOXYLASE AND THE TWO SUBUNITS  
 OF CARBOXYL TRANSFERASE IN A 2:2 COMPLEX (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D10004; BAA00893.1; -;  
 DR EMBL; S77740; AAC60398.1; -;  
 DR EMBL; D63999; BAA10092.1; -;  
 DR PIR; J01238; J01238.  
 DR InterPro; IPR000438; ACOACC\_transf.  
 DR InterPro; IPR000022; Carboxyl\_trans.  
 DR Pfam; PF01039; Carboxyl\_trans; 1.  
 DR PRINTS; PR01070; ACCCTRFRASEB.  
 DR TIGRFAMs; TIGR00515; accd; 1.  
 KW Fatty acid biosynthesis; ligase; zinc-finger; Complete proteome.

FT ZN FING 36 58 C4-TYPE (POTENTIAL).  
 SQ SEQUENCE 326 AA; 36396 MW; 99177806671A858 CRC64;

Query Match 14.7%; Score 359.5; DB 1; Length 326;  
 Best local similarity 36.4%; Pred. No. 1.8e-14;  
 Matches 80; Conservative 43; Mismatches 86; Indels 11; Gaps 2;

QY 44 QGVATLRRASKKACDESVYTGEGTGVAVVLTSPFLGSLGTVASRIMKAIHR 103  
 Db 98 KSYKDRIDPTQANDLTDAVQTHGRLDGLPLAGVMPFRFGSGWSGVGEKLCRLLEY 157  
 QY 104 ATELPLPLVSPASGARMQEDNRAFMVMSITAAVQRRREAHLPFLVYLNPPTGGAMA 163  
 Db 158 ATLERLPVVIICASGARMQEGMSLMQMATISGALQNHREKLYIVLTHPTGGVTA 217  
 QY 164 SWSSGHLTFAEPQAQIFLGPVVELTTHALPDGVOQAENLVKTGVIDGIVSPLQRA 223  
 Db 218 SFAMLGDLILAEPRATIGFAGRRVIEQTLREKLPDDFTSEVILHGHVDAIVAPQLKR 277  
 QY 224 AVAKTKYIQVENVTRDPSPTGVALPVMFAIRSDPQ 263  
 Db 278 TLAQLISLHQP-----FYP-----ILPPLNADSNQVNP 306

## RESULT 7

ACCD\_SOYEN STANDARD; PRT; 432 AA.  
 ID ACCD\_SOYEN  
 AC P49158;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta  
 DE (EC 6.4.1.2) (ACCase beta chain).  
 GN ACCD.  
 OS Glycine max (Soybean).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
 CX NCBI\_TaxID=3847;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=cv. Resnik; TISSUE=leaf;  
 RC Reverdatto S.V., Beilinson V., Nielsen N.C.;  
 RA "The psbL, accD, psal, ORF 203, ORF 151, ORF 103, ORF 229 and psal  
 RT gene cluster in the chloroplast genome of soybean."  
 RL (in) Plant Gene Register FGR95-051.  
 CC -1- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE  
 INVOLVED IN CHLOROPLAST CARBON METABOLISM  
 CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate  
 + malonyl-CoA.  
 CC -1- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U26948; AAA80643.1; -;  
 DR PIR; T06341; T06341.  
 DR InterPro; IPR000438; ACOACC\_transf.  
 DR InterPro; IPR000022; Carboxyl\_trans.  
 DR Pfam; PF01039; Carboxyl\_trans; 1.  
 DR PRINTS; PR01070; ACCCTRFRASEB.  
 DR TIGRFAMs; TIGR00515; accd; 1.  
 KW Fatty acid biosynthesis; ligase; zinc-finger; Chloroplast.  
 FT ZN FING 169 191 C4-TYPE (POTENTIAL).  
 SQ SEQUENCE 432 AA; 48857 MW; 484256A7496ABDD CRC64;

Query Match 13.6%; Score 332.5; DB 1; Length 432;



Best Local Similarity 39.3%; Pred. No. 9 7e-13;  
Matches 81; Conservative 32; Mismatches 82; Indels 11; Gaps 4;

QY 25 IDPDSFISNNETPOYDNLNGYAETLERARSKAKDESVITGEVIGIPVAVILDSF 84  
DB 222 IDPIEFHS-EEBP-----YKDRIDSYQRTKGLTEAVQTGTGGLNIPVALGMDPQF 272  
QY 85 LGGSGTASVASVIMKAIRHATELKLPLIVSPASGARMQEDNRARVMMVSTIAVQHRRE 144  
DB 273 WGGSGASVAGEKITRLVEYATNQLPLILVCSGARMQEGSLDIOMAKISSALYDYQK 332  
QY 145 AHLPLVYLRLNPMTGAGAMASWSSGHLTFAPGAQIGFIPRVVELTTGHALPDGVOA 203  
DB 333 NKKLVYSILTSPTTGGVTAFTGMDITIAEPNAYIAPAGKRVIEQTLNKAIVEGSOQA 392  
QY 204 ENLVKTGVIDGIVSPLOLRAVAATKL 229  
DB 393 EYLFGHGLFDSIV-PRNLKGVLSL 417

## RESULT 8

ACCD\_BRANA STANDARD; PRT; 489 AA.

AC P48937;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta  
DE (EC 6.4.1.2) (ACCase beta chain).  
GN ACCD.

OS Brassica napus (Rape).  
OC Chloroplast

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucots II; Brassicales; Brassicaceae; Brassica.

NCBI\_TaxID=3708;  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Jet neuf;  
RX MEDLINE=96207569; PubMed=8670092;

RA Elborough K.M., Wanz R., Deke R.K., Markham J.E., White A.J.,  
RA Rawsthorne S., Slabas A.R.;

RT "Biotin carboxyl carrier protein and carboxyltransferase subunits of  
RT the multi-subunit form of acetyl-CoA carboxylase from Brassica napus:  
RT cloning and analysis of expression during oilseed rape

embryogenesis.";

RL Biochem. J. 315:103-112(1996).

CC -1- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE  
CC INVOLVED IN CHLOROPLAST CARBON METABOLISM.

CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate  
CC + malonyl-CoA.

CC -1- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; 250868; CA90747.1; -

DR PIR; S66564; S66564.

DR InterPro; IPR000438; ACOACC.transf.

DR InterPro; IPR000022; Carboxyl.trans.

DR Pfam; PF01039; Carboxyl.trans.1.

DR PRINTS; PRO1070; ACCCTFRASEB.

DR TIGRFAMs; TIGR00515; accd.1.

KW Fatty acid biosynthesis; ligase; zinc-finger; Chloroplast.

FT ZN FING 229 248 C4-TYPE (POTENTIAL).

SO SEQUENCE 489 AA; 55476 MW; 540C79FB2A39432 CRC64;

Query Match 13.5%; Score 331.5; DB 1; Length 489;

Best Local Similarity 34.4%; Pred. No. 1.3e-12;  
Matches 76; Conservative 50; Mismatches 82; Indels 13; Gaps 3;

QY 21 IDSVLPDSFISNN-----ETPOYDNLNGYAETLERARSKAKDESVITGEVIG 71  
DB 260 IELSIDG-----SWNGMEDWVNSADPIKFHREERYKRIASQKGTGLTAIGTQQLN 316  
QY 72 GIPVAVILDSFISGLGVASVIRIMKAIRHATELKLPLIVSPASGARMQEDNRARV 131  
DB 317 GIPVALGVMPQFQWGGSMGVGDKITRLILEYATNQLPLILVCSGARMQEGSLIMQ 376  
QY 132 MVSTIAVQHRRE-HIPLVYLNPMTGAMASWSSGHLTFAPGAQIGFIPRVVEL 190  
DB 377 MAKISVLCDDQSSKLPYLSILTSPTTGGVTAFTGMDITIAEPNAYIAPAGKRVIE 436  
QY 191 TTGHALPDGVOAENLVKTGVIDGIVSPLOLRAVAATKL 231  
DB 437 TLKKAIVEGSOQAMSLRKGLDLAIVPRNPLKGVLSL 477

## RESULT 9

ACCD\_LOTUA STANDARD; PRT; 501 AA.

AC Q9BBS1;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta  
DE (EC 6.4.1.2) (ACCase beta chain).  
GN ACCD.

OS Lotus japonicus.

OC Chloroplast

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucots I; Fabales; Fabaceae; Papilionoideae; Lotaseae; Lotus.

NCBI\_TaxID=34305;  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Accession MG-20;  
RX MEDLINE=21082929; PubMed=11214967;

RA Kato T., Kaneko T., Sato S., Nakamura Y., Tabata S.;

RT "Complete structure of the chloroplast genome of a legume, Lotus  
RT japonicus.";

RL DNA Res. 7:333-330(2000).

CC -1- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE  
CC INVOLVED IN CHLOROPLAST CARBON METABOLISM.

CC -1- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate  
CC + malonyl-CoA.

CC -1- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AP002983; BAB33205.1; -

DR InterPro; IPR000438; ACOACC.transf.

DR InterPro; IPR000022; Carboxyl.trans.

DR Pfam; PF01039; Carboxyl.trans.1.

DR PRINTS; PRO1070; ACCCTFRASEB.

DR TIGRFAMs; TIGR00515; accd.1.

KW Fatty acid biosynthesis; ligase; zinc-finger; Chloroplast.

FT ZN FING 235 257 C4-TYPE (POTENTIAL).

SO SEQUENCE 501 AA; 57038 MW; C08BBB343D7DE54D CRC64;

Query Match 13.5%; Score 330.5; DB 1; Length 501;  
Best Local Similarity 36.3%; Pred. No. 1.5e-12;  
Matches 77; Conservative 42; Mismatches 78; Indels 15; Gaps 4;  
QY 21 IDSVLPDSFISNN-----ETPOYDNLNGYAETLERARSKAKDESVITGEVIG 71

```

Db      269 IELSLIDPG---TWNPEDEDMISVDPIEFHSEEPYKDRIDSQKTTGLTEAVOTGCHLN 325
Qy      72 GIPVAVIISDFSLGSLGTVASVRIMKAIHRATELKLPLVSPASGARMOEDNRAFMN 131
Db      336 GIPVAVIIGIMDFEFGMSGWSVGEKILTRLEVAITNQLPLIIVCASGARMOEGSLSLMQ 385
Qy      132 NWSITAAVQRRH-EAHLPLVLYLRNPTMGAMASMGSGHLFPAPGAOIGFAPRVETL 190
Db      386 MAKISSALYDYLANKKLFFVYSILTSPTTGATRASFGMLGDIITIAEPNATIAFRKRVIEQ 445
Qy      191 TTGHALPDGVQOAEMLVKTGVIDGIY--SPLQ 220
Db      446 TLNKAIVBEGSQAAEYLFHKGLFDSIVRNPPLK 477

RESULT 10
ACCD_FORPU STANDARD; PRT; 288 AA.
ID ACCD_FORPU
AC P51198;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
   (EC 6.4.1.2) (ACCase beta chain).
GN ACCD.
OS Porphyra purpurea.
OC Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
OX NCBI_TaxId=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Avonport;
RA Reith M.E., Munholand J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
   genome."
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC - FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
   INVOLVED IN CHLOROPLAST CARBON METABOLISM.
CC - CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
   + malonyl-CoA.
CC - SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, U38804; AAC08084.1; -.
DR PIR, S73119; S73119.
DR InterPro, IPR000438; ACOACC.transFB.
DR InterPro, IPR000022; Carboxyl.trans.
DR Pfam, PF01039; Carboxyl.trans. 1.
DR PRINTS, PRO1070; ACCCTRFRASEB.
DR TIGRfam, TIGR00515; accd; 1.
KW Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.
FT ZN FING 34
FT SEQUENCE 288 AA; 32168 MW; C2FDD4C8A04D8E261 CRC64;

Query Match 13.4%; Score 329; DB 1; Length 288;
Best Local Similarity 38.6%; Pred. No. 9.7e-13;
Matches 71; Conservative 39; Mismatches 74; Indels 0; Gaps 0;

```

```

Qy      166 GSSGHLTPAEPGAOIGFAPRVETLTGHALPDGVQOAEMLVKTGVIDGIAPLRAAV 225
Db      218 AMLGDLIIAEPKALIAFAGRRVIEOTIKEDLPDNFOSSEYLFHFGPLDILIVERTQRSKL 277
Qy      226 AKTL 229
Db      278 IQIL 281

RESULT 11
ACCD_TOBAC STANDARD; PRT; 512 AA.
ID ACCD_TOBAC
AC P12219;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
   (EC 6.4.1.2) (ACCase beta chain).
GN ACCD OR YCP11 OR ZPPA.
OS Nicotiana tabacum (Common tobacco).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxId=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Bright yellow 4;
RA Shinozaki K., Ohme M., Tanaka M., Wakaugi T., Hayashida N.,
RA Matsubayashi T., Zaita N., Chunwongse J., Obokata J.,
RA Yamaguchi-Shinozaki K., Ohno C., Torazawa K., Meng B.-Y., Sugita M.,
RA Dene H., Kamogashira T., Yamada K., Kusuda J., Takiwa F., Kato A.,
RA Tohoh N., Shimada H., Sugita M.;
RT "The complete nucleotide sequence of the tobacco chloroplast genome:
   its gene organization and expression."
RL EMBO J. 5:2043-2049(1986).
CC - FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE
   INVOLVED IN CHLOROPLAST CARBON METABOLISM.
CC - CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate
   + malonyl-CoA.
CC - SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, Z00044; CAAT7362.1; -.
DR PIR, A05196; A05196.
DR InterPro, IPR000438; ACOACC.transFB.
DR InterPro, IPR000022; Carboxyl.trans.
DR Pfam, PF01039; Carboxyl.trans. 1.
DR PRINTS, PRO1070; ACCCTRFRASEB.
DR TIGRfam, TIGR00515; accd; 1.
KW Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.
FT ZN FING 247
FT SEQUENCE 512 AA; 58472 MW; DFEC312E6228868 CRC64;

Query Match 13.3%; Score 326.5; DB 1; Length 512;
Best Local Similarity 34.4%; Pred. No. 2.7e-12;
Matches 75; Conservative 49; Mismatches 87; Indels 7; Gaps 2;

```

```

Qy      46 YAEATERARSKACBESVITGEGTVEGIPVAVIISDFSLGSLGTVASVRIMKAIHRYAT 105
Db      98 YSQRKUDAFKTSLDQAVOTGVGTMLGQKVCIGINDFRMGSGMSGVGEKILTRLEKRT 157
Qy      106 ELKPLVSPASGARMOEDNRAFMNWSITAAVQRRHRAHLPLVLYLRNPTMGAMASW 165
Db      158 QERLPAILICASGARMOEGSLSLMQMAKISSALEMKKENLLYLSVLTSPTTGVTASIF 217

```

```

Qy      21 IDVSVDPSFSSWNETP-----QYDNLNGQYAEITLSEARSKACBESVITGEGTVEGIP 74
Db      261 IELSLIDPGTWNPEDEDMISVDPIEFHSEEPYKDRIDSQKTTGLTEAVOTGCHLN 340
Qy      75 VAVIISDFSLGSLGTVASVRIMKAIHRATELKLPLVSPASGARMOEDNRAFMNWS 134
Db      341 VALGVMDFQFMGSGMSGVGEKILTRLEVAITNQLPLIIVCASGARMOEGSLSLMQAK 400

```



Db 389 ISSAFDYOENLVTGVIDGIVSPLOLRAVAKTLKV 231  
 194 HALPDGOAENLVTGVIDGIVSPLOLRAVAKTLKV 231  
 449 KIVPESQMAEYLFQKGLFDLIIIPNLLKSLVGLSEFL 486

## RESULT 14

ACCD ANGLY STANDARD; PRT; 314 AA.  
 ID ACCD ANGLY  
 AC P28252;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta  
 (EC 6.4.1.2) (ACCase beta chain).  
 GN ACCD OR YCF11 OR ZFPA OR DEDB.  
 OS Angiopteris lygodilfolia (Turnip fern).  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Filicophyta; Marattiopsida; Marattiales; Marattiaceae; Angiopteris.  
 OX NCBI\_Taxid=3267;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=rosensack;  
 RX MEDLINE=92119238; PubMed=1731980;  
 RA Yoshinaga K., Kubota Y., Ishii T., Wada K.;  
 RT "Nucleotide sequence of atpB, tbcL, trnR, deeb and psal chloroplast  
 genes from a fern Angiopteris lygodilfolia: a possible emergence of  
 Spermatophyta lineage before the separation of Bryophyta and  
 Pteridophyta.";  
 RL Plant Mol. Biol. 18:79-82(1992).  
 CC -!- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE  
 INVOLVED IN CHLOROPLAST CARBON METABOLISM.  
 CC -!- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate  
 + malonyl-CoA.  
 CC -!- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; X58429; CAA41333.1; -.  
 DR PIR; S19230; BWFNZT.  
 DR InterPro; IPR000438; ACOACC\_transfB.  
 DR InterPro; IPR000022; Carboxyl\_trans.  
 DR Pfam; PF01039; Carboxyl\_trans\_1.  
 DR PRINTS; PR01070; ACCCTRFRASEB.  
 DR TIGRFAMs; TIGR00515; accd; 1.  
 DR TrEMBL; TIGR00515; accd; 1.  
 KW Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.  
 FT ZN FING 51 73 C4-TYPE (POTENTIAL).  
 SQ SEQUENCE 314 AA; 35098 MW; 0809A7DF118DD28 CRC64;

Query Match 13.2%; Score 322; DB 1; Length 314;  
 Best Local Similarity 34.1%; Pred. No. 2.8e-12;  
 Matches 74; Conservative 46; Mismatches 91; Indels 6; Gaps 2;

Qy 21 IDSVLDPSPFSIMNETPOYDNL-----NCGVAFLETERARSKAKCDESVITGGTVEGIPV 75  
 Db 85 IELLDRIWIPMDDMVAQDVLAFTSDEDSYNNRISLSQKRGTLDAVQTGINANGPVP 144  
 Qy 76 AVILSDPFLGSLCTVASVRIMKAIRHATELKLPLVSPASGARMOEDNRAFYMMVTSI 135  
 Db 145 ALGVMDPFGSMGSMGVGKTRILEVATQESLPLIIVCASGARMEGTLSSQMOMAKI 204  
 Qy 136 TAAVORHR-EAHLPLVTLRNPTMGAMASWGSSGHLTFAEQAQIGFLGPVVELTTGGH 194  
 Db 205 SSVSQIHQVOKKLTAVLAVLTPTTGGVTASFGMLDIIIAEPKAVIAFAGKRVIBQTLRQ 264

Qy 195 ALPDGOAENLVTGVIDGIVSPLOLRAVAKTLKV 231  
 Db 265 KIPDFQVAESLFDHGLDSIVPRLLKGVLSERTEL 301

## RESULT 15

ACCD SPIOL STANDARD; PRT; 522 AA.  
 ID ACCD SPIOL  
 AC 09M3L7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta  
 (EC 6.4.1.2) (ACCase beta chain).  
 GN ACCD.  
 OS Spinacia oleracea (Spinach).  
 OG Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
 OX NCBI\_Taxid=3562;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Geant d'hiver, and cv. Monatol;  
 RX MEDLINE=21187424; PubMed=11292076;  
 RA Schmitz-Lineweber C., Maier R.M., Alcaraz J.-P., Cottet A.,  
 RA Hermann R.G., Maché R.;  
 RT "The plastid chromosome of spinach (Spinacia oleracea): complete  
 nucleotide sequence and gene organization.";  
 RL Plant Mol. Biol. 45:307-315(2001).  
 CC -!- FUNCTION: PROBABLE SUBUNIT OF A BIOTIN-DEPENDENT CARBOXYLASE  
 INVOLVED IN CHLOROPLAST CARBON METABOLISM.  
 CC -!- CATALYTIC ACTIVITY: ATP + acetyl-CoA + HCO(3) (-) = ADP + phosphate  
 + malonyl-CoA.  
 CC -!- SIMILARITY: BELONGS TO THE ACCD / PCCB FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; AJ400848; CAB88738.1; -.  
 DR InterPro; IPR000438; ACOACC\_transfB.  
 DR InterPro; IPR000022; Carboxyl\_trans.  
 DR Pfam; PF01039; Carboxyl\_trans\_1.  
 DR PRINTS; PR01070; ACCCTRFRASEB.  
 DR TIGRFAMs; TIGR00515; accd; 1.  
 DR TrEMBL; TIGR00515; accd; 1.  
 KW Fatty acid biosynthesis; Ligase; Zinc-finger; Chloroplast.  
 FT ZN FING 258 280 C4-TYPE (POTENTIAL).  
 SQ SEQUENCE 522 AA; 58907 MW; 84A337F76A3BF87 CRC64;

Query Match 13.1%; Score 321.5; DB 1; Length 522;  
 Best Local Similarity 34.1%; Pred. No. 5.3e-12;  
 Matches 76; Conservative 45; Mismatches 85; Indels 17; Gaps 4;

Qy 21 IDSVLDPSPFSIMNETPOYDNL-----NCGVAFLETERARSKAKCDESVITGGTVEGIPV 69  
 Db 292 IELLDIPG---TWN--PMDDDMVSMNDPIGFHSEEBAYDRIDSQIKRGLEAVQTGIGQ 346  
 Qy 70 VEGIPVAVILSDPFLGSLCTVASVRIMKAIRHATELKLPLVSPASGARMOEDNRAFY 129  
 Db 347 INGIPVAIGVMDPFGSMGSMGVGKTRILEVATQESLPLIIVCASGARMEGTLSSQMOMAKI 406  
 Qy 130 VMMVSTAAVORHREAHLPVLY-VLRNPTMGAMASWGSSGHLTFAEQAQIGFLGPVVEL 188  
 Db 407 MQMKAISVLYDYOSNKKLFTVSLTSTPTGGVTASFGMLDIIIAEPNAYIAFAGKRVIV 466  
 Qy 189 ELTTGHALPDGOAENLVTGVIDGIVSPLOLRAVAKTLKV 231

Db 467 EQLINKTVPEGSOAAFLFHKGLEDPDIVPRNLKGVLSELPFL 509

Search completed: November 13, 2003, 06:41:00  
Job time : 27 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 05:43:33 ; Search time 89 Seconds

(without alignments)  
1423.639 Million cell updates/sec

Title: US-10-024-370-3

Perfect score: 2448

Sequence: 1 VEKREPTVMWGMHTSALT.....NNPERAGDRSFRFRRLAQ 491

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapexc 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2445	99.9	491	16	Q93KF1 corynebacte
2	1527.5	62.4	483	16	Q8FR58 corynebacte
3	599.5	24.5	578	2	Q9F6D6 streptomyc
4	480	19.6	458	16	Q9L0A1 streptomyc
5	431.5	17.6	301	16	Q8Y5E8 brucella me
6	431.5	17.6	301	16	Q8FXV7 brucella su
7	427	17.4	308	16	Q8CNS5 rhizobium
8	409	16.7	1338	5	Q9GQ08 giardia lam
9	404	16.5	272	2	Q8G6D2 heliobacilli
10	392.5	16.0	316	16	Q8YU12 anabaena sp
11	389.5	15.9	304	16	Q92TC7 rhizobium m
12	387.5	15.8	289	16	Q92K01 heliobacte
13	386	15.8	309	16	Q8DIE7 heliobacte
14	384	15.7	289	16	Q8DIE7 heliobacte
15	379	15.5	201	2	Q50132 mycobacteri
16	375	15.3	304	16	Q8XFS5 salmonella

17	374	15.3	332	16	Q8FFH5 escherichia
18	372.5	15.2	284	16	Q9RV15 deinococcus
19	370.5	15.1	282	16	Q9K841 bacillus ha
20	369.5	15.1	276	16	Q66754 aquifex ao
21	368	15.0	288	16	Q9CHP2 lactococcus
22	367	15.0	304	16	Q8RGA0 fusobacteri
23	365	14.9	291	16	Q8E1K1 streptococc
24	365	14.9	291	16	Q8E1K1 streptococc
25	363	14.8	262	16	Q34571 bacillus su
26	361.5	14.8	298	16	Q8U7A8 agrobacteri
27	361.5	14.8	308	16	Q84295 chlamydia t
28	361.5	14.8	308	16	Q929C3 chlamydia p
29	361	14.7	283	16	Q9PKA1 chlamydia m
30	360.5	14.7	288	16	Q8D5X0 streptococc
31	359	14.7	290	16	Q8XSV3 ralstonia s
32	355.5	14.5	280	16	Q9P100 campylobac
33	354.5	14.5	285	16	Q8CS66 staphylococ
34	346.5	14.2	290	16	Q9K0D3 neisseria m
35	345.5	14.1	290	16	Q9JVD9 neisseria m
36	345.5	14.1	304	16	Q8ZD25 yersinia pe
37	345.5	14.1	356	16	Q8D0U1 agrobacteri
38	344	14.1	317	16	Q8U9Y5 agrobacteri
39	343.5	14.0	285	16	Q9T5Z2 staphylococ
40	343.5	14.0	285	16	Q8NW59 staphylococ
41	343.5	14.0	304	16	Q9CN12 pasteurella
42	343	14.0	288	16	Q8NZN5 streptococc
43	343	14.0	288	16	Q8K6J3 streptococc
44	339	13.8	297	8	Q8W9X6 chaetosphe
45	338.5	13.8	295	16	Q8PJ30 xanthomonas

## ALIGNMENTS

RESULT 1	ID	Q93KF1	PRELIMINARY;	PRT;	491 AA.
AC	Q93KF1	01-DEC-2001 (TRENBLREL. 19, Created)			
DT	01-DEC-2001 (TRENBLREL. 19, Last sequence update)				
DT	01-MAR-2003 (TRENBLREL. 23, Last annotation update)				
DE	Putative carboxyltransferase subunit of acetyl-CoA carboxylase				
DE	(Acetyl-CoA carboxylase beta subunit) (EC 6.4.1.2).				
GN	ACCA OR CGL0831.				
OS	Corynebacterium glutamicum (Brevibacterium flavum).				
OC	Bacteria; Actinobacteria; Actinobacteriales;				
OC	Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.				
OX	NCBI_TaxID=1718;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ATCC 13032;				
RA	Kim K., Peters-Wendisch P.G., Sahn H., Birkmann B.J.;				
RN	Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.				
RM	Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-ATCC 13032 / DSM 20300 / NCIB 10025;				
RA	Nakagawa S.;				
RT	Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.				
RL	EMBL, Y17592; CAC42827.1; -				
DR	EMBL, AP005276; BAB98224.1; -				
DR	InterPro: IPR000438; ACOACC; transfe.				
DR	InterPro: IPR000022; Carboxyl; trans.				
DR	InterPro: IPR001753; EnCoA_hydites.				
DR	Pfam: PF00378; ECH; 1.				
DR	Pfam: PF00378; ECH; 1.				
DR	PRINTS: PR01070; ACCCTPRASEB.				
KM	ligase: Transferrase: Complete proteome.				
SO	SEQUENCE 491 AA; 52238 MW; 53D1B8757A92D82D CRC64;				

Query Match 99.9%; Score 2445; DB 16; Length 491;  
Best Local Similarity 99.8%; Pred. No. 3.5e-138;  
Matches 490; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 VEKRPVWGMERTSALTLLIDSVLPDPSFISWNETPOYDNLNOGYAETLERARAKACD 60
   1 MEKRPVWGMERTSALTLLIDSVLPDPSFISWNETPOYDNLNOGYAETLERARAKACD 60
Db 1 MEKRPVWGMERTSALTLLIDSVLPDPSFISWNETPOYDNLNOGYAETLERARAKACD 60
Qy 61 ESVITGEGTVEGIPVAVILSDPSFLGSLGTVA SVIRKAIHRATELKLPLVSPASGA 120
   61 ESVITGEGTVEGIPVAVILSDPSFLGSLGTVA SVIRKAIHRATELKLPLVSPASGA 120
Db 61 ESVITGEGTVEGIPVAVILSDPSFLGSLGTVA SVIRKAIHRATELKLPLVSPASGA 120
Qy 121 RMOENRNFVMMVSTTA VORRREAHLPVLVLRNPTMGAMASWGSSGHLTFAPGQOI 180
   121 RMOENRNFVMMVSTTA VORRREAHLPVLVLRNPTMGAMASWGSSGHLTFAPGQOI 180
Db 121 RMOENRNFVMMVSTTA VORRREAHLPVLVLRNPTMGAMASWGSSGHLTFAPGQOI 180
Qy 181 GFELGRVVELTTGHALPDGVOAENLVKTGVLDIGVSPLOLRAAVALKVIQPEATDR 240
   181 GFELGRVVELTTGHALPDGVOAENLVKTGVLDIGVSPLOLRAAVALKVIQPEATDR 240
Db 181 GFELGRVVELTTGHALPDGVOAENLVKTGVLDIGVSPLOLRAAVALKVIQPEATDR 240
Qy 241 FSPPTPGVALPVMEAIARSRDPORPGIEIMETLGADVVKLSGARAGALSPAVRVALARI 300
   241 FSPPTPGVALPVMEAIARSRDPORPGIEIMETLGADVVKLSGARAGALSPAVRVALARI 300
Db 241 FSPPTPGVALPVMEAIARSRDPORPGIEIMETLGADVVKLSGARAGALSPAVRVALARI 300
Qy 301 GGRPVVILIGODRRFTLGPOELRFAARGISLARELMPLVSIIDTSGAEISQAABELGIAS 360
   301 GGRPVVILIGODRRFTLGPOELRFAARGISLARELMPLVSIIDTSGAEISQAABELGIAS 360
Db 301 GGRPVVILIGODRRFTLGPOELRFAARGISLARELMPLVSIIDTSGAEISQAABELGIAS 360
Qy 361 SIARTLSKLIDAPLPTVSVITIGGVGGALAMLPA DLVYAAENAMLSALPPEGASAILFR 420
   361 SIARTLSKLIDAPLPTVSVITIGGVGGALAMLPA DLVYAAENAMLSALPPEGASAILFR 420
Db 361 SIARTLSKLIDAPLPTVSVITIGGVGGALAMLPA DLVYAAENAMLSALPPEGASAILFR 420
Qy 421 DTNHAABIIRERGVOAHALLSGGLIDGIVAE TEHFVEEILGTISNALSELDDNNPERAGRD 480
   421 DTNHAABIIRERGVOAHALLSGGLIDGIVAE TEHFVEEILGTISNALSELDDNNPERAGRD 480
Db 421 DTNHAABIIRERGVOAHALLSGGLIDGIVAE TEHFVEEILGTISNALSELDDNNPERAGRD 480
Qy 481 SRFTEFERLAQ 491
   481 SRFTEFERLAQ 491
Db 481 SRFTEFERLAQ 491

```

## RESULT 2

```

Q8FR58 PRELIMINARY; PRT; 483 AA.
ID Q8FR58
AC Q8FR58;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Putative carboxyltransferase subunit of acetyl-CoA carboxylase.
GN ACCDA OR CE0907.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Ueda Y., Sugimoto S.,
RT "The entire genomic sequence of Corynebacterium efficiens YS-314."
RL Submitted (MAY-2002) to the EMBL/Genbank/DBD databases.
DR EMBL; AP005217; BAC17717.1; -
KW Transferase; Complete proteome.
SQ SEQUENCE 483 AA; 51545 MW; 653BA89519BBDC7 CRC64;

```

Query Match 62.4%; Score 1527.5; DB 16; Length 483;  
 Best Local Similarity 62.9%; Pred. No. 2.2e-83;  
 Matches 302; Conservative 62; Mismatches 115; Indels 1; Gaps 1;

```

Qy 12 MEHSALTLLIDSVLPDPSFISWNETPOYDNLNOGYAETLERARAKACDESITGEGTVE 71
   12 MEHSALTLLIDSVLPDPSFISWNETPOYDNLNOGYAETLERARAKACDESITGEGTVE 71
Db 1 MTRTSARDLIDVLDKDSFCFMDTPPEYGEIDDDIRALARRKRSQVDEAVITGEGTVE 60
   1 MTRTSARDLIDVLDKDSFCFMDTPPEYGEIDDDIRALARRKRSQVDEAVITGEGTVE 60
Qy 72 GIPVAVILSDPSFLGSLGTVA SVIRKAIHRATELKLPLVSPASGARMQEDRRARVFM 131
   72 GIPVAVILSDPSFLGSLGTVA SVIRKAIHRATELKLPLVSPASGARMQEDRRARVFM 131

```

```

Db 61 GSRVAFIVSEFAFLGSGIGATSRRLLIOAIRATRERLPLLISPCSGTRMOEGTAPAFAM 120
Qy 132 MVSTTA VORRREAHLPVLVLRNPTMGAMASWGSSGHLTFAPGQOIGFLGRVVELT 191
   132 MVSTTA VORRREAHLPVLVLRNPTMGAMASWGSSGHLTFAPGQOIGFLGRVVELT 191
Db 121 MISTITAVTRKMDHLPVLVLRNPTMGAMASWGSSGHLTFAPGQOIGFLGRVVELT 180
   121 MISTITAVTRKMDHLPVLVLRNPTMGAMASWGSSGHLTFAPGQOIGFLGRVVELT 180
Qy 192 TGHALPDGVOAENLVKTGVLDIGVSPLOLRAAVALKVIQPEATDRSPPTPGVALP 251
   192 TGHALPDGVOAENLVKTGVLDIGVSPLOLRAAVALKVIQPEATDRSPPTPGVALP 251
Db 181 TGTPEBGVQRGENTLAHGVIDGVSPLOLRAAVALKVIQPEATDRSPPTPGVALP 240
   181 TGTPEBGVQRGENTLAHGVIDGVSPLOLRAAVALKVIQPEATDRSPPTPGVALP 240
Qy 252 VMEAIARSRDPORPGIEIMETLGADVVKLSGARAGALSPAVRVALARIGRPVVLIGOD 311
   252 VMEAIARSRDPORPGIEIMETLGADVVKLSGARAGALSPAVRVALARIGRPVVLIGOD 311
Db 241 AMESIMITREGRGGRGVADLVSRILGRNVVTLSGTGDDGISATVVALARIGRPVVLIGOD 300
   241 AMESIMITREGRGGRGVADLVSRILGRNVVTLSGTGDDGISATVVALARIGRPVVLIGOD 300
Qy 312 R-RTLGPOELRFAARGISLARELMPLVSIIDTSGAEISQAABELGIASIAITLSKLI 370
   312 R-RTLGPOELRFAARGISLARELMPLVSIIDTSGAEISQAABELGIASIAITLSKLI 370
Db 301 RHLPLGPALRRFRRRIALAREQLPISIIDTSGAEISQAABELGIASIAITLSKLI 360
   301 RHLPLGPALRRFRRRIALAREQLPISIIDTSGAEISQAABELGIASIAITLSKLI 360
Qy 371 DAPLPTVSVITIGGVGGALAMLPA DLVYAAENAMLSALPPEGASAILFRDTNHAABIE 430
   371 DAPLPTVSVITIGGVGGALAMLPA DLVYAAENAMLSALPPEGASAILFRDTNHAABIE 430
Db 361 DAPLPTVSVITIGGVGGALAMLPA DLVYAAENAMLSALPPEGASAILFRDTNHAABIE 420
   361 DAPLPTVSVITIGGVGGALAMLPA DLVYAAENAMLSALPPEGASAILFRDTNHAABIE 420
Qy 431 RQGVQAHALLSGGLIDGIVAE TEHFVEEILGTISNALSELDDNNPERAGRSRFTFERLA 490
   431 RQGVQAHALLSGGLIDGIVAE TEHFVEEILGTISNALSELDDNNPERAGRSRFTFERLA 490
Db 421 RQGVSAHAMLASGIIDGLIAEHDDFGEVVAIDNTLHELTMBERVGRGQRFHYERLA 480
   421 RQGVSAHAMLASGIIDGLIAEHDDFGEVVAIDNTLHELTMBERVGRGQRFHYERLA 480

```

## RESULT 3

```

Q9F6D6 PRELIMINARY; PRT; 578 AA.
ID Q9F6D6
AC Q9F6D6;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Acetyl carboxylase.
GN ZHOF.
OS Streptomyces sp. R1128.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=140437;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1128;
RC MEDLINE=20517899; PubMed=10931852;
RA Marti T., Hu Z., Pohl N.L., Shah A.N., Khosla C.;
RT "Cloning, Nucleotide Sequence, and Heterologous Expression of the
RT Biosynthetic Gene Cluster for R1128, a Non-steroidal Estrogen Receptor
RT Antagonist. Insights into an Unusual Priming Mechanism."
RL J. Biol. Chem. 275:33443-33448(2000).
DR EMBL; AF293442; AAC30193.1; -
DR InterPro; IPR001095; Ac-CoA carboxylase.
DR InterPro; IPR000438; ACOACC_transf.
DR InterPro; IPR000022; Carboxyl_trans.
DR Pfam; PF03255; ACCA; 1.
DR Pfam; PF01039; Carboxyl_trans; 1.
DR PRINTS; PR01069; ACCCTFRASEA.
DR PRINTS; PR01070; ACCCTFRASEB.
DR TIGRFAWS; TIGR00515; accd; 1.
SQ SEQUENCE 578 AA; 60667 MW; 979330136912C323 CRC64;

```

Query Match 24.5%; Score 599.5; DB 2; Length 578;  
 Best Local Similarity 34.2%; Pred. No. 7.9e-28;  
 Matches 175; Conservative 71; Mismatches 218; Indels 47; Gaps 9;

```

Qy 21 IDSVLPDPSFISWNETPOYDNLNOGYAET-----LERARAKACDESITGEGTVECI 73
   21 IDSVLPDPSFISWNETPOYDNLNOGYAET-----LERARAKACDESITGEGTVECI 73
Db 52 LDQLDPSGSAEPRIINDVTRFDVL--GFTDTPYPARLAARAATGLDEAVAVITGRIHQ 109
   52 LDQLDPSGSAEPRIINDVTRFDVL--GFTDTPYPARLAARAATGLDEAVAVITGRIHQ 109
Qy 74 PVAVILSDPSFLGSLGTVA SVIRKAIHRATELKLPLVSPASGARMQEDRRARVFMV 133
   74 PVAVILSDPSFLGSLGTVA SVIRKAIHRATELKLPLVSPASGARMQEDRRARVFMV 133
Db 110 PLVAVMDFAFLGSLGAVGELLTTLAETALAEHLPLVSPASGARMQEDRRARVFMV 169
   110 PLVAVMDFAFLGSLGAVGELLTTLAETALAEHLPLVSPASGARMQEDRRARVFMV 169
Qy 134 SITAAVORRREAHLPVLVLRNPTMGAMASWGSSGHLTFAPGQOIGFLGRVVELTTG 193
   134 SITAAVORRREAHLPVLVLRNPTMGAMASWGSSGHLTFAPGQOIGFLGRVVELTTG 193

```



```

Db      170 KTSALGQDEAGVLTISLVPTVGVVAAFPATLCLVIAEPGALGPAGRVIAQTIR 229
      194 HALPGVOQAEMLVKTVGDIIVSLQLEAAVAKTLK-----10PEVATDRFST 244
      220 BELPPEFOTAEYLHSHGLIDLIVPRRLDELARLRACAKPKGGDGPARGGLDACA 289
      245 TPGVALPV-----MEAIARSRDQRPQIGEMETLGVKLSGARAGALSPAVRVAL 297
      290 SPYVVDPADLAHNDPSRQVBARAIDRPTLDVIALCLDFELKQDRISGDCPAIVAGL 349
      298 ARIGRPVVLIGOD-----RRFTLG---PQELRPARGISLARELNLPIVSIIDTSG 346
      350 ARFGGRPVAIVIGHQKHDAQOELRRRNFQKPIPSGYRKARIMRLADKGLPVTFVDTG 409
      347 AELSGAAEELGIASSIARTLSKLIADPLPTVSIIIGCGVGGALMLPDLVYAENML 406
      410 AYPGAVAEBQGAVALAENLRIMAGLRVVTVTGEGSGALALAVANRVLMEENGTY 469
      407 SALPEGASAILFRDPTNHAAILIERQGVQAHLLSGGLDGIIVATE-----HFVEE 458
      470 SVISEGCAILIMQPPAAGRAARALRYTARELLGLGIVDSVLPPEGGTGADPLRADA 529
      459 ILGTISNALISE---LDNNPERAGDSRPTRF 486
      530 LRDAIAGLAELSLPDGHQLVADRWARFRRY 560

```

## RESULT 4

Q9LOAI PRELIMINARY; PRT; 458 AA.

```

AC 09LOAI;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Putative acetyl CoA carboxylase (Alpha and beta subunits).
GN SC02445 OR SCC24.16.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
CN NCBI_TaxID=1902;
RN 01
RC SEQUENCE FROM N.A.
RA Brown S.P., Harris D.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN 02
RC SEQUENCE FROM N.A.
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN 03
RC SEQUENCE FROM N.A.
RA MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinaishi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN 04
RC SEQUENCE FROM N.A.
RA STRAIN=A3(2) / M145;
RA MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Jablonowski E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Metcortrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;

```

```

RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RT Nature 417:141-147(2002).
DR EMBL, AL939112, CAB86110.1, -.
DR InterPro, IPR000438, AC0ACC_transfB.
DR InterPro, IPR000022, Carboxyl_trans.
DR Pfam, PF01039, Carboxyl_trans, 1.
DR PRINTS, PR01070, ACCCTRFRASEB.
KM Complete proteome.
SQ SEQUENCE 458 AA; 47180 MW; E5E4086CB29B89C1 CRC64;

```

Query Match 19.6%; Score 480; DB 16; Length 458;  
 Best Local Similarity 31.1%; Pred. No. 7.9e-21;  
 Matches 138; Conservative 57; Mismatches 193; Indels 56; Gaps 9;

```

      16 SALTLIDSLVLDPDSTISNNEPQYDNLNOGIAETLERARSAKCDSEYITGEYEGIPV 75
      21 SELPHADPWRBPDGLG-----PGVDASPARAAERTGEAESVVCCTGLVEGTRA 70
      76 AVILDSFSLGSGITVASVRIMKAIHRATELKLPLVSPASGARMDENRAPFMVMSI 135
      71 VLVSTFEFGLGSLGHRGDRLEAAVAYAREHRLFPVPLVATGSRMOBMLALTOLQRV 130
      136 TAAVQRHREAHLPFLVLRNPTMGAMASWSSGHLTFAEPGAQIGFLAPRVBELTQHA 195
      131 ARQSALTRAAAGLAQIAVVRDPATGGWATLTAGADVVALALGAGVFGAGSRV-----R 183
      196 LPD---GVQAEMLVKTGIVDIGIVSLQLRPAVAKTLKVI-----QPEVATRF 241
      184 PPDADPAAYTAEQAQVAGSADAVVPPGELRATLGRWMLTAPSGCPAPVEPELGARD-- 241
      242 SPTTPGVALPV--MEAIARSRDQRPQIGEMETLGVKLSGARAGALSP-ANRVALA 298
      242 -----LPAGMEAVARRADPGPRAGAVTIDAVFTRVALISGRCGRDPEGMLCGPG 293
      299 RIGGRPVVLIGQDRRFTLGPELRFARRGISLARELNLPIVSIIDTSGAELSGAAEELGI 358
      294 SHAGRTVAVAAQTGAT-RRAGVYRTATLHLBARLGLPVLTVDTPPAADAEMERAGA 352
      359 ASSIARTLSKLIADPLPTVSIIIGCGVGGALMLPDLVYAENAMLSPPEGASAIL 418
      353 GPATADLFCAVASARTPTVTIVIGEGSGGALALAPGSTWATPDSYSLAIE----- 406
      419 FRDTHAAELIERQGVQAHLLSQ 442
      407 -----HAAAILKRPPEAVEATAGQ 425

```

## RESULT 5

O8YE58 PRELIMINARY; PRT; 301 AA.

```

AC O8YE58;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
DE (EC 6.4.1.2).
GN BME12020.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
CN NCBI_TaxID=29459;
RN 01
RC SEQUENCE FROM N.A.
RA STRAIN=16M / ATCC 23456 / Biotype 1;
RA MEDLINE=20020109; PubMed=1175668;
RA DelVecchio V.G., Kaparatou I., Redkar R.J., Patra G., Mujer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonowski L., Larsen N., D'Souza M., Bernal A., Nazur M., Goldsman E.,
RA Selkov E., Elizer P.H., Hagius S., O'Callaghan D., Teleson J.-D.,
RA Hagelcorn R., Kyrides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";

```

RESULT 8  
Q9GQ08  
ID Q9GQ08 PRELIMINARY; PRT; 1338 AA

AC Q9GQ08; 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)  
 DE Putative acetyl-CoA carboxylase/pyruvate carboxylase fusion protein.  
 OS Giardia lamblia (Giardia intestinalis).  
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.  
 RN NCBI\_TaxID=5741;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WB/ATCC30957;  
 RA Hense K., Muller M.;  
 RT "A putative Acetyl-CoA Carboxylase/Pyruvate Carboxylase fusion protein in the amoebocyte protist, Giardia lamblia.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF207740; AAG35712.1; -  
 DR InterPro; IPR001095; Ac-CoA carboxylase.  
 DR InterPro; IPR000438; ACoACC\_transf.  
 DR InterPro; IPR000089; Biotin lipoyl.  
 DR InterPro; IPR000022; Carboxyl\_trans.  
 DR InterPro; IPR000891; HMG1-like.  
 DR InterPro; IPR003379; PYC\_OXA.  
 DR Pfam; PF00364; biotin lipoyl; 1.  
 DR Pfam; PF01039; Carboxyl\_trans; 1.  
 DR Pfam; PF02436; PYC\_OXA; 1.  
 DR PRINTS; PRO1059; ACCTRRPASA.  
 DR PRINTS; PRO1070; ACCCTRRPASEB.  
 DR Pyruvate.  
 SQ SEQUENCE 1338 AA; 148221 MW; 8DB0125EA5FAED1 CRC64;

Query Match 16.7%; Score 409; DB 5; Length 1338;  
 Best Local Similarity 27.5%; Pred. No. 6.1e-16;  
 Matches 132; Conservative 90; Mismatches 192; Indels 66; Gaps 15;

QY 30 FISMNETPOYDNLNQG--YAETLERAR--SKAKDESVITGEGEIPVAVILDSFPL 85  
 DB 9 FINLSTHMDVFFGVGSYKQLEAARAKTSLCMHYHVEPEYIRA--VQEFAM 65  
 QY 86 GSGSLGTVAVRIMKAIHRATELKLPLVSPASGARMQEDNRAFYVMVSTAAVRHBA 145  
 DB 66 GSGSLGTVAVRIMKAIHRATELKLPLVSPASGARMQEDNRAFYVMVSTAAVRHBA 145  
 QY 146 HLPFLVLRNPTMGAMASWSSGSHLTFAEPGAQIGLCPRVVELTT-----GHALP 197  
 DB 126 GIPVLVLRNPTMGAMASWSSGSHLTFAEPGAQIGLCPRVVELTT-----GHALP 197  
 QY 198 DGVOQAEMLVKTGVLDGIVSPLO--LRAAVAKTLKVI-----QPEATD-----RFSPTT 245  
 DB 186 AGFQTTDRAAATGALDQLPDSLESAYASIGHILEVLIQKMQSVBAAKTALRLTGV 245  
 QY 246 PGVALPVMEAI-----ARSRDPRPGIGRIMETLGDVVKLSGABAGALSPAVRYAL 297  
 DB 246 HGAEDVTEDAFVQDAFYRDAENVTRPQCDYLKMF-----VMAASLARDKCIITAF 299  
 QY 298 ARIGER-FVVLIGORRFTLQGE-----LRFRRGISLARENLPIVSIITDSGABL 349  
 DB 300 GVFDLTLVLIATSRREDVLPDEAGLSPTGYRFAARAIYLANRLSIPVITLVDAGALP 359  
 QY 350 SQAAEELGIASSIARTLSKLIDAPLPVSVIIIGGVGGALAMLPADIVYAENMGL 409  
 DB 360 SPEAEEOGOSRAISCLAFQAVNHPITSIITGEGSGGALLAGNVTGILSKAFVYVI 419  
 QY 410 PPEGASAIL-----FDTNHAARI-----IEROGVQVANA-----LLSQGLIDGIVAEER 454  
 DB 420 SPEGASAILGFSVYR--NNTAEMRANFISDAELAHVQRCVPIDRLNNGVNVNDIIVGEY 477

RESULT 9  
 Q8GDQ2 PRELIMINARY; PRT; 272 AA.  
 AC Q8GDQ2;  
 AC Q8GDQ2;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, last annotation update)  
 DE Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta (EC 6.4.1.2) (Fragment).  
 OS Hellobacillus mobilis.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Hellobacteriaceae;  
 RN NCBI\_TaxID=28064;  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=22337798; PubMed=12446909;  
 RA Raymond J., Zhaxybayeva O., Gogarten J.P., Gerdes S.Y.,  
 RA Blankenship R.E.;  
 RT "Whole-genome analysis of photosynthetic prokaryotes.";  
 RL Science 298:1616-1620(2002).  
 DR EMBL; AY142924; AAN87528.1; -  
 DR Transferase; Ligase.  
 KW NON TER 272  
 FT NON TER 272  
 SQ SEQUENCE 272 AA; 29796 MW; 721D9DA2BA18812D CRC64;

Query Match 16.5%; Score 404; DB 2; Length 272;  
 Best Local Similarity 41.5%; Pred. No. 1.3e-16;  
 Matches 90; Conservative 40; Mismatches 85; Indels 2; Gaps 2;

QY 21 IDSVDPSPFISWN-ETPOYNLN--QGYAETLERARSKAKDESVITGEGEIPVAVI 78  
 DB 50 IDSLIDEGTLEMDLVTKDPLNPPGYAKIQKQSTMAEALITGQGIHGPVAVIG 109  
 QY 79 LSDSFLLGSLGTVAVRIMKAIHRATELKLPLVSPASGARMQEDNRAFYVMVSTAA 138  
 DB 110 AMDSRFIMASGAVVGEKIVAAEKLEKMPILIFSASGARQEGVLSIMQARTSAA 169  
 QY 139 VQHRREAHLPVLVLRNPTMGAMASWSSGSHLTFAEPGAQIGLCPRVVELTTGHALPD 198  
 DB 170 LTRLEAGLPFVSFVLTPTTGGVTFASFMALDILLAEAGALIGFTGPRVIRQTRKPLA 229  
 QY 199 GVOQAEMLVKTGVLDGIVSPLO--LRAAVAKTLKVI-----QPEATD-----RFSPTT 245  
 DB 230 GFQSEFLQKHGWVDIILERPMMKRAQLATLLSLHMPM 266

## RESULT 10

QY 08YU12 PRELIMINARY; PRT; 316 AA.  
 AC 08YU12;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)  
 DE Acetyl-CoA carboxylase beta subunit.  
 GN ACCD OR ALU2364.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 RN NCBI\_TaxID=103690;  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21595285; PubMed=11759840;  
 RA Kaneke T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsumoto A., Muraki A.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AP003589; BAB74063.1; -  
 DR InterPro; IPR000438; ACoACC\_transf.  
 DR InterPro; IPR000022; Carboxyl\_trans.

DR Pfam; PF01039; Carboxyl\_trans; 1.  
DR PRINTS; PRO1070; ACCCTPRASEB.  
DR TIGRFAMs; TIGR00515; accd; 1.  
KW Complete proteome.  
SQ SEQUENCE 316 AA; 35455 MW; FDB192F20BDDF633E CRC64;

Query Match	16.0%;	Score 392.5;	DB 16;	Length 316;
Best Local Similarity	38.0%;	Pred. No. 7.9e-16;		
Matches 87;	Conservative 42;	Mismatches 81;	Indels 19;	Gaps 3;

```

QY 32 SNAETFOYDUL-----NOGVAETTERARSXAKDESIYITGEGBVBIIPAIVLSD 81
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85 TWR--PMDEULRATDPLQFRODRKAYSIDRLAREMEDTLGLIDAVKTCGLQINSFALAVMD 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 82 FSELGSLGTVAWSVRIMKAIHRAETELKPLVLPASGGARMOEDNRAFVMWVSITAAVOR 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 143 FRFEGSGMSVGEKITRLEIQOTRRYPPVILITSGSGARMOEGMLSLMQAKISALER 202
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 142 HREHLEPLVYLKPNPIMGAMASWSSGHLTAAEFQAQIGFGRPVVLTITGHALLPQGVQ 201
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 HRARLLIYIEVLNPTPTGGVITASFAMLAGDIILAEKPAITGFGRVRIEQTREKLPDDFQ 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 202 QAEULVYTGVIDGIWBPLOLRAAVAKTLKVIOPVEATBRFSPTPGVAL 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 263 TAEDLLHGFVDIVIPRPTQKNTLSQJLIHOPV-----PTTPBWL 304
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 11

```

ID      092ATC7      PRELIMINARY;
AD      092ATC7;
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      Probable acetyl-Coenzyme A carboxylase subunit
DE      beta protein (EC 6.4.1.2).
GN      ACCD OR R00030 OR SMC02764.
OS      Rhizobium meliloti (Sinorhizobium meliloti).
OC      Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC      Rhizobiaceae; Sinorhizobiolum.
OX      NCBI_TaxId=382;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=1021;
RX      MEDLINE=21396507; PubMed=11481430;
RA      Capela D., Barloy-Hubler F., Gouzy J., Boche G., Ampe F., Batut J.,
RA      Boissard P., Becker A., Boucny M., Cadieu E., Diano S., Gloux S.,
RA      Godrie T., Goffeau A., Kahn D., Kiss E., Lelaune V., Masny D.,
RA      Pohl T., Portetelle D., Puchet A., Purnelle B., Rambergier U.,
RA      Renard C., Thebault P., Vandenbol M., Weidner S., Gallbert F.,
RT      "Analysis of the chromosome sequence of the legume symbiont
RT      Sinorhizobium meliloti strain 1021."
RL      Sinorhizobium meliloti strain 1021."
DR      EMBL; AL591782; SCA4117.1; -.
DR      ProC. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR      InterPro; IPR000438; AC0ACX
DR      InterPro; IPR000022; AC0ACXyl trans.
DR      Pfam; PF01039; Carboxyl trans. 1.
DR      PRINTS; PR01070; ACCCRFRASEB.
DR      TIGRFAMs; TIGR00515; acccd; 1.
DR      LIGase; Complete proteome.
SQ      SEQUENCE 304 AA; 33567 MW; 4FCC45B674D068AB CRC64;

```

Query Match	15.9%	Score 389.5;	DB 16;	Length 304;
Best Local Similarity	39.2%	Pred. No. 1.1e-15;		
Matches 82;	Conservative 44;	Mismatches 80;	Indels 3;	Gaps 2;

[illegible]

```
QY      166 GSSCHLTFAEPGAOIGFLGRVVELTTGHALPDGVQQAEMLVKTGVIDIGTSPQLRAAV 225
        | : ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
DB     213 AMLCDIHMAERGAETGFACRIVIEDTLREKLPREGQTSEYLEHGNVDMMVKRHDIPEYL 272
```

```

QY      226 AKTLKVI-QPVEATDRFSPTTGGVALPV 252
      ||::|||
DB      273 ARVLNIIMKKPAKAVKR-DTATELAPLPV 300

```

## RESULT 12

ID	09ZKOJ1	PRELIMINARY;	PRT;	289 AA.
AD	09ZKOJ1			
DT	01-MAY-1999	(TEMBLrel. 10, Created)		
DT	01-MAY-1999	(TEMBLrel. 10, Last sequence update)		
DT	01-JUN-2002	(TEMBLrel. 21, Last annotation update)		
DE	Acetyl-coenzyme A carboxylase subunit B.			
GN	ACCD OR JHP0884.			
OS	Helicobacter pylori J99 (Campylobacter pylori J99).			
OC	Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;			
OC	Helicobacteraceae; Helicobacter.			
OX	NCBI_TaxId=85963;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99120557; PubMed=9923682;			
RA	Aim R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.			
RA	Smith D.R., Noonan B., Guidi B.C., deJonge B.L., Carmel G.			
RA	Thumma P.J., Caruso A., Uria-Nichelsen M., Mills D.M., Ives C,			
RA	Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F			
RA	Trust T.J.;			
RT	"Genomic sequence comparison of two unrelated isolates of the human			
RT	gastric pathogen Helicobacter pylori.";			
RL	Nature 397:176-180 (1999).			
DR	EMBL; AE001518; AAD06468.1; -.			
DR	InterPro; IPR000438; ACACCC.transfB.			
DR	InterPro; IPR000022; Carboxyl.trans.			
DR	Pfam; PF01039; Carboxyl.trans; 1.			
DR	PRINTS; PR01070; ACCCTRPRASB.			
DR	TIGRFAMs; TIGR00515; accd; 1.			
QW	Complete proteome.			
SEQUENCE	289 AA;	32567 MW;	43B6EAAEB879E630 CRC64;	

Query Match	15.8%;	Score 387.5;	DB 16;	Length 289;
Best Local Similarity	39.6%;	Pred. No. 1.4e-15;		
Matches 91;	Conservative 43;	Mismatches 81;	Indels 15;	Gaps 4

```

QY      10 SATLILSDVLDDPSFSJMNME--TPQYDNLN----QGYAETLERARSJAKDESYITTEGT 69
Dh      60 NATERLEFLCDVGSFEEDPKILRPV-DPLNVDVKESTYQRIKKYEKTTNRPSYSIGEAK 118
QY      70 VEGIPAVILSDPSFLGSGSLGTVASVRIMKAIRHATELTKLPLTVPASGARGMOEDNRAF 129
Dh      119 INMPLQIYVFPFSGFMGSGLSVBESEKIVRAINRAVAKKELLILVNSGARGMOESTYSL 178
QY      130 VMNVSITAAVQSHREAHLPFLVYLRLNPLTMGAMAMASWGSSGHLTAEPQAOIGFLGPRVVE 189
Dh      179 MQMAKTSALNLTLSSEAKLPFLISLSDPTVGGVSAFPAFLGLLIAEPGAMIGFAGPRVIK 238
QY      190 LTTGHAHLPRGVCQAENLVKTVGVIDGTSVPLQLRANA-----KTLKV 231
Dh      239 QITIGADLPFGQTAEFLLEGLIMDIYHAKDOKTTLTLAMMTHTKSI 288

```

RESULT 13  
08DIE7

ID	ORDIS7 AC	PRELIMINARY:	PRI;	309 AA.
DT	01-MAR-2003	(TREMblrel. 23, Created)		
DT	01-MAR-2003	(TREMblrel. 23, Last sequence update)		
DT	01-MAR-2003	(TREMblrel. 23, Last annotation update)		
DE	Acetyl-CoA carboxylase beta subunit.			
DE	ACCD OR TLR1643.			

OS Synechococcus elongatus (Thermosynechococcus elongatus).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
 NCBI\_TaxID=32046;  
 [1]  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BP-1;  
 RX MEDLINE=2225144; PubMed=12240834;  
 RA Nakamura Y., Kaneo T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
 Matanabe A., Iriuchi M., Kawashima K., Kimura T., Kishida Y.,  
 Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A., Nakazaki N.,  
 Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.,  
 RT "Complete genome structure of the thermophilic cyanobacterium  
 Thermosynechococcus elongatus BP-1."  
 RL DNA Res. 9:123-130(2002).  
 DR EMBL; AP005374; BAC09195.1; -.  
 SQ Complete proteome.  
 Query Match 15.8%; Score 386; DB 16; Length 309;  
 Best Local Similarity 37.0%; Pred. No. 1.9e-15;  
 Matches 94; Conservative 46; Mismatches 96; Indels 18; Gaps 5;  
 QY 11 GHEH-TSALTLDVIDPDSFISWNETPQYDNI-----NCGYATLERASKAC 59  
 DB 56 GHHRVSSSERIQQLDPH---TWR--PDENLVSCDPIQFKDKYSDRLREYQKGL 110  
 QY 60 DESVTIGEGTVRGIPAVVILSDPSFGSLGTVASRIMKAIHRATELKLPLVSPASG 119  
 DB 111 KQAVQTGLQQLSPLALGVMDPAFVGSGMSVGKIKTLIRATWEHIPLVIVCASG 170  
 QY 120 ARMQEDNRAFVWVMSITAAVORHREALPLVLYLRNPVTMGAMASGSGHLTPAEPAQ 179  
 DB 171 ARMQSGMSLMGMAKTAALERRHSGLYIPILHTPTMGVYASAMGDIITIAEPKAT 230  
 QY 180 IGFLEPRVELTTGHALPDGVOAENLVKTVGIDGIVSPQLRAAVAKTKLVQPEAVD 239  
 DB 231 IAFAGRRVIEQTLREKLPDFAEFVQCGFVDIVPRTQKSLIARLRIHQVPS-- 288  
 QY 240 RSPPTPGVALPYM 253  
 DB 289 RMSALLPKSFPLM 302

RESULT 14  
 025604 PRELIMINARY; PRT; 289 AA.  
 AC 025604;  
 DT 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Acetyl-CoA carboxylase beta subunit (ACCD).  
 GN HP0950.  
 OS Helicobacter pylori (Campylobacter pylori).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
 OC Helicobacteraceae; Helicobacter.  
 NCBI\_TaxID=210;  
 [1]  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=97394467; PubMed=9252185;  
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
 Fletschmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,  
 Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 Berg D.E., Gocayne J.D., Uiterback T.R., Peterson J.D., Kelley J.M.,  
 Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wattey L., Wallin E.,  
 Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Frazer C.M.,  
 Venter J.C.;  
 RT "The complete genome sequence of the gastric pathogen Helicobacter  
 pylori."  
 RL Nature 388:539-547(1997).  
 DR EMBL; AE00604; AAD07995.1; -.

DR TIGR; HP0950; -.  
 DR InterPro; IPR000438; ACOACC\_transf.  
 DR InterPro; IPR000022; Carboxyl\_trans.  
 DR Pfam; PF01039; Carboxyl\_trans; 1.  
 DR PRINTS; PRO1070; ACCCTRRPASEB.  
 DR TIGRFAMs; TIGR00515; accd; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 289 AA; 32553 MW; A57ADDAF3AB0AFC CRC64;  
 Query Match 15.7%; Score 384; DB 16; Length 289;  
 Best Local Similarity 39.3%; Pred. No. 2.3e-15;  
 Matches 84; Conservative 40; Mismatches 72; Indels 18; Gaps 2;  
 QY 26 DDPSPFISWNETPQYONLNGYATLERASKACDESVTIGEGTVRGIPAVVILSDPSFL 85  
 DB 85 DPLNFDV-----KSYKRIKTKYKRTNRPSVSGAKTNRPQLIVDFDSFM 134  
 QY 86 GSGLTGVSVRIMKAIHRATELKLPLVSPASGARMOEDNRAFYWVMSITAAVORHREA 145  
 DB 135 GSGLSGVEGEKIVRAINNAVAREALLIVSASGARMGESTVSLQMATSAALRLSEA 194  
 QY 146 HLPFLVLYLRNPVTMGAMASGSGHLTPAEPAQIGFLEPRVELTTGHALPDGVOAEN 205  
 DB 195 KLPFTSLSDPYGVSASFALGDLIIAEPGAMIGFAPRVIKOTIGADLPFGFOTAEF 254  
 QY 206 LKTVGIDGIVSPQLRAVA-----KTLKV 231  
 DB 255 LIEHGLIDVIVARKDKTSLDILMMTHTKSKI 288

RESULT 15  
 050132 PRELIMINARY; PRT; 201 AA.  
 AC 050132;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE znfa.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NCBI\_TaxID=1769;  
 [1]  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Smith D.R.;  
 RL Submitted (Apr-1995) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Robison K.;  
 RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; U15184; AAA63045.1; -.  
 DR InterPro; IPR000438; ACOACC\_transf.  
 DR InterPro; IPR000022; Carboxyl\_trans.  
 DR Pfam; PF01039; Carboxyl\_trans; 1.  
 DR PRINTS; PRO1070; ACCCTRRPASEB.  
 SQ SEQUENCE 201 AA; 21323 MW; EF45B8DCBB804FB4 CRC64;  
 Query Match 15.5%; Score 379; DB 2; Length 201;  
 Best Local Similarity 44.4%; Pred. No. 2.7e-15;  
 Matches 87; Conservative 28; Mismatches 73; Indels 8; Gaps 3;  
 QY 122 MOEDNRAFVWVMSITAAVORHREALPLVLYLRNPVTMGAMASGSGHLTPAEPAQIG 181  
 DB 1 MOEGTVAFLQWVKIAAAVAKLSKRAGLPYLGILRNPTTGGVSAWSGLGHTVAEPGLIG 60  
 QY 182 FLEPRVELTTGHALPDGVOAENLVKTVGIDGIVSPQLRAAVAKTKLVQPEAVTRF 241  
 DB 61 FLGPRVLYLKGPRPSIQTARENLRGIVDAIVTLDGLTLNRALTMTADV---PKL 117  
 QY 242 SPT---TPGVALPVEAIARSQDPQRPDGEIMETTGADVVKLSGARAGALSPAVRYAL 297  
 DB 118 IPTQRPPEIPDVDAWNSVMSGRPERPSVAQVLRHGATDVLVLSGPGHGAATTL-LAL 176

Qy	298	ARIGRPVVLIGDDR	313
		: : : :	
Db	177	ARLAGTAVVIGQQRK	192

Search completed: November 13, 2003, 06:42:46  
Job time : 94 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 06:41:05 / Search time 29 Seconds  
(without alignments)  
716.367 Million cell updates/sec

Title: US-10-024-370-3

Perfect score: 2448

Sequence: 1 VEKRPPTVMGMEHTSALT.....NNPERAGRDSRFRERLAQ 491

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/2/1aa/6C\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/2/1aa/6D\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2448	100.0	491	4 US-09-362-899-3	Sequence 3, Appl1
2	413	16.9	215	4 US-08-311-731A-410	Sequence 410, App
3	354.5	14.5	293	4 US-09-134-001C-3803	Sequence 3803, Ap
4	354	14.5	289	4 US-09-107-532A-3824	Sequence 3824, Ap
5	331	13.5	300	4 US-09-328-352-5240	Sequence 5240, Ap
6	328.5	13.5	339	4 US-09-252-991A-28751	Sequence 28751, A
7	319.5	13.1	255	4 US-09-198-452A-76	Sequence 76, Appl
8	289	11.8	317	2 US-08-973-375-1	Sequence 1, Appl1
9	286	11.7	279	4 US-09-328-352-8167	Sequence 8167, Ap
10	283.5	11.6	315	4 US-09-134-001C-3786	Sequence 3786, Ap
11	271.5	11.1	262	4 US-09-107-532A-4837	Sequence 4837, Ap
12	265.5	10.8	344	4 US-09-198-452A-433	Sequence 433, App
13	265.5	10.3	327	4 US-08-611-107-12	Sequence 12, Appl
14	252.5	10.3	327	2 US-08-422-560A-12	Sequence 12, Appl
15	252.5	10.3	327	2 US-08-468-793-12	Sequence 12, Appl
16	250	10.2	332	4 US-09-252-991A-21222	Sequence 21222, A
17	215.5	8.8	543	2 US-08-776-597A-2	Sequence 2, Appl1
18	215.5	8.8	543	2 US-08-693-228-2	Sequence 2, Appl1
19	210	8.6	546	4 US-09-328-352-5584	Sequence 5584, Ap
20	198	8.1	544	4 US-09-252-991A-26096	Sequence 26096, A
21	192	7.8	1116	4 US-09-252-991A-24374	Sequence 24374, A
22	186.5	7.6	78	4 US-08-311-731A-371	Sequence 371, App
23	178	7.3	692	4 US-09-252-991A-27997	Sequence 27997, A
24	177.5	7.3	538	4 US-09-328-352-6530	Sequence 6530, Ap
25	176.5	7.2	566	4 US-09-107-532A-5786	Sequence 5786, Ap
26	155.5	6.4	332	4 US-09-328-352-6531	Sequence 6531, Ap
27	144	5.9	336	4 US-09-252-991A-16707	Sequence 16707, A

28	143.5	5.9	5215	3 US-09-105-537-2	Sequence 2, Appl1
29	128.5	5.2	391	4 US-09-252-991A-24230	Sequence 24230, A
30	126.5	5.2	2254	4 US-08-677-010-3	Sequence 3, Appl1
31	126.5	5.2	2254	2 US-08-790-519-3	Sequence 3, Appl1
32	122	5.0	51	1 US-08-418-893D-4	Sequence 4, Appl1
33	122	5.0	120	4 US-08-311-731A-411	Sequence 411, App
34	122	5.0	3201	4 US-09-679-279-15	Sequence 15, Appl
35	121	4.9	3739	3 US-09-320-878-2	Sequence 2, Appl1
36	121	4.9	3739	3 US-09-105-537-33	Sequence 33, Appl
37	121	4.9	3739	4 US-09-141-908-3	Sequence 3, Appl1
38	121	4.9	3739	4 US-09-657-440-2	Sequence 2, Appl1
39	121	4.9	11877	3 US-09-105-537-6	Sequence 6, Appl1
40	119.5	4.9	841	3 US-09-413-814-107	Sequence 107, App
41	119.5	4.9	4545	2 US-08-804-227C-14	Sequence 14, Appl
42	119.5	4.9	4550	2 US-08-804-227C-8	Sequence 8, Appl1
43	119.5	4.9	4550	2 US-08-804-198-2	Sequence 2, Appl1
44	118	4.8	822	4 US-09-252-991A-23090	Sequence 23090, A
45	118	4.8	2628	3 US-09-413-814-11	Sequence 11, Appl

# ALIGNMENTS

RESULT 1									
US-09-362-899-3									
Sequence 3, Application US/09362899									
Patent No. 6361986									
GENERAL INFORMATION:									
APPLICANT: Degussa-Hls AG									
TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY FERMENTATION AND									
CURRENT APPLICATION NUMBER: US/09/362,899									
CURRENT FILING DATE: 1999-07-29									
EARLIER FILING DATE: 1999-05-27									
NUMBER OF SEQ ID NOS: 3									
SOFTWARE: PatentIn Ver. 2.1									
SEQ ID NO 3									
LENGTH: 491									
TYPE: PRT									
ORGANISM: Corynebacterium glutamicum									
US-09-362-899-3									
Query Match									
Best Local Similarity 100.0%; Score 2448; DB 4; Length 491;									
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	VEKRPPTVMGMEHTSALT	1	IDSVDPSFTSMNETPOYDNLNOGYAETLEBRARSKAKCD	60				
DB	1	VEKRPPTVMGMEHTSALT	1	IDSVDPSFTSMNETPOYDNLNOGYAETLEBRARSKAKCD	60				
QY	61	ESVITGECTGEGIPAVIL	61	SDPSFLGSLGTVAIVIRIKAIHRA TELKPLLVSPASGA	120				
DB	61	ESVITGECTGEGIPAVIL	61	SDPSFLGSLGTVAIVIRIKAIHRA TELKPLLVSPASGA	120				
QY	121	RMQENRFAVMWVSTTAA	121	VQHRREHLPVLYLNPTMGMASSGSGHITFAPPGQOI	180				
DB	121	RMQENRFAVMWVSTTAA	121	VQHRREHLPVLYLNPTMGMASSGSGHITFAPPGQOI	180				
QY	181	GFLGRVVELTTGHALP	181	PGVQAEMLVYTGVIDGIVSPLOQRAAVAKTLKYOPEATDR	240				
DB	181	GFLGRVVELTTGHALP	181	PGVQAEMLVYTGVIDGIVSPLOQRAAVAKTLKYOPEATDR	240				
QY	241	FSPFTPGVALPVMAL	241	ARSRDPORGIGELINETLGADVYKLSGARAGALSPAVRVALARI	300				
DB	241	FSPFTPGVALPVMAL	241	ARSRDPORGIGELINETLGADVYKLSGARAGALSPAVRVALARI	300				
QY	301	GGRPVVILGDRRTTIG	301	QELRPARRGISLARETNPVSIIDPSGALSGAAEBELGIAS	360				
DB	301	GGRPVVILGDRRTTIG	301	QELRPARRGISLARETNPVSIIDPSGALSGAAEBELGIAS	360				
QY	361	STARTSLIDAPPTVSV	361	IIGQGVGGALAMLPAADVAAENMWLSLPEGASATIFR	420				
DB	361	STARTSLIDAPPTVSV	361	IIGQGVGGALAMLPAADVAAENMWLSLPEGASATIFR	420				



Db 361 STARTLSKLIDAPLPTVSIIIGGVGGALAMLPA DLVYAAENAWLSALPEGASAILFR 420  
QY 421 DTMHAEIIRERGVQAHALLSGGLDGYAEFEHEVEELGTISNAISLDPNPERARD 480  
Db 421 DTMHAEIIRERGVQAHALLSGGLDGYAEFEHEVEELGTISNAISLDPNPERARD 480  
QY 481 SRFTFRERLAQ 491  
Db 481 SRFTFRERLAQ 491

RESULT 2  
US-08-311-731A-410  
; Sequence 410, Application US/08311731A  
; Patent No. 6583266  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, DOUGLAS  
; APPLICANT: MAO, JEN-I  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 411  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
; STREET: 600 ATLANTIC AVENUE  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/311,731A  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GATES, EDWARD R.  
; REGISTRATION NUMBER: 31,616  
; REFERENCE/DOCKET NUMBER: C0044/7125  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/720-3500  
; TELEFAX: 617/720-2441  
; INFORMATION FOR SEQ ID NO: 410:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 215 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Mycobacterium leprae  
; US-08-311-731A-410

Query Match 16.9%; Score 413; DB 4; Length 215;  
Best Local Similarity 45.6%; Pred. No. 3.9e-34;  
Matches 94; Conservative 29; Mismatches 75; Indels 8; Gaps 3;

QY 112 LVSPASGARMQEDRAFMVMSITAAVORREAHLPFLVYLIRNPTMGAMASGSGHL 171  
Db 5 LVSPSSDGTMRQEGVAFLOWKTAAYKLSKRAGLPLGLIRNPTTGGVSPASGSLGHV 64  
QY 172 TPAEBAQIGFLGPRVELTTGHALPDGVQAEMLVKTGVIDGIVSPQLRAAAVAKTLKV 231  
Db 65 TVAEFGVIGFLGPRVELLYGEPSPSDIQTAENLORHGVDAIVTLDTGLQTLIRALTM 124  
QY 232 IQPVEATDRFSPPT---TPGVALPMEALIASRDPORPGIGEMETTLGADVYKLSGARAG 287  
Db 125 IADV---PKLPTQRPSPIDVPVPMNSVMSRRRERSVAQVLRHGATDVLLSGPHG 181  
QY 288 ALSPAVVALARIIGRPVVLIGQDR 313

Db 182 EAATTL-LALARLAGQTAIVTIGQOR 206

RESULT 3  
US-09-134-001C-3803  
; Sequence 3803, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3803  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
; US-09-134-001C-3803

Query Match 14.5%; Score 354.5; DB 4; Length 293;  
Best Local Similarity 36.0%; Pred. No. 6.4e-28;  
Matches 82; Conservative 48; Mismatches 93; Indels 5; Gaps 3;

QY 13 EHTSLTL---IDSLDPSFISWNETPOYDN-LN-OGVAETLERAKSKAKDESVITGE 67  
Db 64 DHIHALTRAKRIEASISBGSFIERDRGMTSANLPDPFGYEKEIKQOQKGLAEALVSGT 123  
QY 68 GTVEGIPVAVILSDPSFLGSLGTVASVRIMKAIHRATELKLPLVSPASGARMQEDNR 127  
Db 124 AKLDGIQYVAVMDARFRMGSMGVSVEKICRIIDYCTEHRPLPILSASGARMQEGII 183  
QY 128 AFVMMVSTAAVORREAHLPFLVYLIRNPTMGAMASGSGHLTPAEBAQIGFLGPRV 187  
Db 184 SLMGKTSVSLKSHSAGLGLYSIYNPTTGGVSASFASVGDINSEPRALIGFAGRRV 243  
QY 188 VELTTGHALPDGVQAEMLVKTGVIDGIVSPQLRAAAVAKTLKV 235  
Db 244 IEOTINKLPDDPDTAFLEHGLDKVIRKQMRRTLSNIIKHQEV 291

RESULT 4  
US-09-107-532A-3824  
; Sequence 3824, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598

```
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Atinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3824:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...289
; SEQUENCE DESCRIPTION: SEQ ID NO: 3824:
US-09-107-532A-3824

Query Match          14.5%: Score 354; DB 4; Length 289;
Best Local Similarity 35.5%: Pred. No. 7e-28;
Matches 77; Conservative 45; Mismatches 79; Indels 16; Gaps 2;

QY 24 VLDPDSFISMET-----POYDNLNOGYAETLERARSKACDESVITGEGVEGIP 74
DB 70 IIDEKSFEMPDLDVLTOKPLAFPE-----YTEKIKKQDGTGHEAVLTGKATIGIIP 122
QY 75 VAVILSDPSFLGSGIGTASVRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVS 134
DB 123 FLIGMDNFMTGMSGITVIGETITLPERATKESLPLVLTFTASGARMQEGIFSLMQAK 182
QY 135 ITAAVORREARLPLVYLRLNPTMGAMASGSGHLTFAPBGAQIGFLPRVVELTTGH 194
DB 183 ISAAVKRNSNGLFLVLTPTTGTGVTNSFMESGDIIIAEPQSILGPRGRVVEIQTIKQ 242
QY 195 ALPDGVOQAEMLVKTGVIDIGIVSPLOLRRAVAKTLKV 231
DB 243 ELPEDFQKAEFLSHGFPVDQIVPRMELKQKHTLLEL 279

RESULT 5
US-09-328-352-5240
; Sequence 5240, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5240
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5240

Query Match          13.5%: Score 331; DB 4; Length 300;
Best Local Similarity 36.6%: Pred. No. 1.7e-25;
Matches 75; Conservative 37; Mismatches 83; Indels 10; Gaps 1;

QY 23 SVLPDPSISNRETQYDNLNOGYAETLERARSKACDESVITGEGVEGIPVAVILSDF 82
DB 97 SAKPDLKLV-----SKPYDPKREKQTKGTETALAMOGNINGVMDIMACAEF 146
QY 83 SFLGSLGTASVRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVSITAAVGRH 142
```

```
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Atinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3824:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...289
; SEQUENCE DESCRIPTION: SEQ ID NO: 3824:
US-09-107-532A-3824

Query Match          14.5%: Score 354; DB 4; Length 289;
Best Local Similarity 35.5%: Pred. No. 7e-28;
Matches 77; Conservative 45; Mismatches 79; Indels 16; Gaps 2;

QY 24 VLDPDSFISMET-----POYDNLNOGYAETLERARSKACDESVITGEGVEGIP 74
DB 70 IIDEKSFEMPDLDVLTOKPLAFPE-----YTEKIKKQDGTGHEAVLTGKATIGIIP 122
QY 75 VAVILSDPSFLGSGIGTASVRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVS 134
DB 123 FLIGMDNFMTGMSGITVIGETITLPERATKESLPLVLTFTASGARMQEGIFSLMQAK 182
QY 135 ITAAVORREARLPLVYLRLNPTMGAMASGSGHLTFAPBGAQIGFLPRVVELTTGH 194
DB 183 ISAAVKRNSNGLFLVLTPTTGTGVTNSFMESGDIIIAEPQSILGPRGRVVEIQTIKQ 242
QY 195 ALPDGVOQAEMLVKTGVIDIGIVSPLOLRRAVAKTLKV 231
DB 243 ELPEDFQKAEFLSHGFPVDQIVPRMELKQKHTLLEL 279

RESULT 5
US-09-328-352-5240
; Sequence 5240, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5240
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5240

Query Match          13.5%: Score 331; DB 4; Length 300;
Best Local Similarity 36.6%: Pred. No. 1.7e-25;
Matches 75; Conservative 37; Mismatches 83; Indels 10; Gaps 1;

QY 23 SVLPDPSISNRETQYDNLNOGYAETLERARSKACDESVITGEGVEGIPVAVILSDF 82
DB 97 SAKPDLKLV-----SKPYDPKREKQTKGTETALAMOGNINGVMDIMACAEF 146
QY 83 SFLGSLGTASVRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVSITAAVGRH 142
```

```
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Atinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3824:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...289
; SEQUENCE DESCRIPTION: SEQ ID NO: 3824:
US-09-107-532A-3824

Query Match          14.5%: Score 354; DB 4; Length 289;
Best Local Similarity 35.5%: Pred. No. 7e-28;
Matches 77; Conservative 45; Mismatches 79; Indels 16; Gaps 2;

QY 24 VLDPDSFISMET-----POYDNLNOGYAETLERARSKACDESVITGEGVEGIP 74
DB 70 IIDEKSFEMPDLDVLTOKPLAFPE-----YTEKIKKQDGTGHEAVLTGKATIGIIP 122
QY 75 VAVILSDPSFLGSGIGTASVRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVS 134
DB 123 FLIGMDNFMTGMSGITVIGETITLPERATKESLPLVLTFTASGARMQEGIFSLMQAK 182
QY 135 ITAAVORREARLPLVYLRLNPTMGAMASGSGHLTFAPBGAQIGFLPRVVELTTGH 194
DB 183 ISAAVKRNSNGLFLVLTPTTGTGVTNSFMESGDIIIAEPQSILGPRGRVVEIQTIKQ 242
QY 195 ALPDGVOQAEMLVKTGVIDIGIVSPLOLRRAVAKTLKV 231
DB 243 ELPEDFQKAEFLSHGFPVDQIVPRMELKQKHTLLEL 279

RESULT 5
US-09-328-352-5240
; Sequence 5240, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5240
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5240

Query Match          13.5%: Score 331; DB 4; Length 300;
Best Local Similarity 36.6%: Pred. No. 1.7e-25;
Matches 75; Conservative 37; Mismatches 83; Indels 10; Gaps 1;

QY 23 SVLPDPSISNRETQYDNLNOGYAETLERARSKACDESVITGEGVEGIPVAVILSDF 82
DB 97 SAKPDLKLV-----SKPYDPKREKQTKGTETALAMOGNINGVMDIMACAEF 146
QY 83 SFLGSLGTASVRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVSITAAVGRH 142
```

```
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Atinello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3824:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...289
; SEQUENCE DESCRIPTION: SEQ ID NO: 3824:
US-09-107-532A-3824

Query Match          14.5%: Score 354; DB 4; Length 289;
Best Local Similarity 35.5%: Pred. No. 7e-28;
Matches 77; Conservative 45; Mismatches 79; Indels 16; Gaps 2;

QY 24 VLDPDSFISMET-----POYDNLNOGYAETLERARSKACDESVITGEGVEGIP 74
DB 70 IIDEKSFEMPDLDVLTOKPLAFPE-----YTEKIKKQDGTGHEAVLTGKATIGIIP 122
QY 75 VAVILSDPSFLGSGIGTASVRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVS 134
DB 123 FLIGMDNFMTGMSGITVIGETITLPERATKESLPLVLTFTASGARMQEGIFSLMQAK 182
QY 135 ITAAVORREARLPLVYLRLNPTMGAMASGSGHLTFAPBGAQIGFLPRVVELTTGH 194
DB 183 ISAAVKRNSNGLFLVLTPTTGTGVTNSFMESGDIIIAEPQSILGPRGRVVEIQTIKQ 242
QY 195 ALPDGVOQAEMLVKTGVIDIGIVSPLOLRRAVAKTLKV 231
DB 243 ELPEDFQKAEFLSHGFPVDQIVPRMELKQKHTLLEL 279

RESULT 5
US-09-328-352-5240
; Sequence 5240, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5240
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5240

Query Match          13.5%: Score 331; DB 4; Length 300;
Best Local Similarity 36.6%: Pred. No. 1.7e-25;
Matches 75; Conservative 37; Mismatches 83; Indels 10; Gaps 1;

QY 23 SVLPDPSISNRETQYDNLNOGYAETLERARSKACDESVITGEGVEGIPVAVILSDF 82
DB 97 SAKPDLKLV-----SKPYDPKREKQTKGTETALAMOGNINGVMDIMACAEF 146
QY 83 SFLGSLGTASVRIMKAIHRATELKLPLVSPASGARMQEDNRAVMMVSITAAVGRH 142
```

NAME/KEY: SITE  
LOCATION: 1...255  
OTHER INFORMATION: Xaa=unknown or other  
US-09-198-452A-76

Query Match 13.1%; Score 319.5; DB 4; Length 255;  
Best Local Similarity 38.7%; Pred. No. 2e-24;  
Matches 75; Conservative 30; Mismatches 74; Indels 15; Gaps 3;

16 SALTIDVDPDSTISMETPOYDNL-----NQGAFETIERARSAKCEBSYIT 65  
DB 60 TALEVVKLADDD---SWR--PLYDLKSGQDLEFIDTDTYANRLEKARKTTSEGVIV 114  
QY 66 GEGVEGIPVAVILSDFSFLGSLGTVASVRIMKAIRHATELKLPLVSPASGARMQED 125  
DB 115 GICITGLHVALAVNDFFMAGSMKAVGXKLTILIEERLIERLVIIVISASGARMQES 174  
QY 126 NRAFVMMVSTDAVORHREAHLPVLVYLRNPTMGAMASWSSGHLTPAEPGAQIGFLGP 185  
DB 175 VESLMQWKTSAALAKLHEAGLPYISVLTNPISGVTSPALGDIIIAEPKALICFAPGP 234  
QY 186 RYVELTTHGHALPDG 199  
DB 235 RVVAQVIGEDLPEG 248

## RESULT 8

US-08-973-275-1  
Sequence 1, Application US/08973275B  
Patent No. 5958706

GENERAL INFORMATION:  
APPLICANT: MATSUNAGA, Tadashi  
APPLICANT: KAMIYA, Shinji  
TITLE OF INVENTION: FINE MAGNETIC PARTICLES CONTAINING USEFUL PROTEINS  
TITLE OF INVENTION: BOUND THEREOF, PROCESS FOR PRODUCING THE SAME, AND USE  
FILE REFERENCE: MATSUNAGA  
CURRENT APPLICATION NUMBER: US/08/973,275B  
EARLIER FILING DATE: 1998-02-09  
EARLIER APPLICATION NUMBER: PCT/JP97/01043  
EARLIER FILING DATE: 1997-03-27  
EARLIER APPLICATION NUMBER: JP 8-97536  
EARLIER FILING DATE: 1996-03-28  
EARLIER APPLICATION NUMBER: JP 8-146833  
EARLIER FILING DATE: 1996-05-16  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 317  
TYPE: PRT  
ORGANISM: Magnetospirillum AMB-1  
US-08-973-275-1

Query Match 11.8%; Score 289; DB 2; Length 317;  
Best Local Similarity 29.2%; Pred. No. 3.8e-21;  
Matches 100; Conservative 56; Mismatches 130; Indels 56; Gaps 10;

171 LTFAPGAQIGFLGPRVVELTTHGHALPDG--VQQAENLVKGVIDGIVSPLOJRAVAKT 228  
DB 4 LEFERPIAE---LEGKIEL---RHLSDBGDVNIDAVSK-----LQAKVKL 45  
QY 229 LKVIQPEVETDFSPITTPVALPWMEAKRSRDPQPGIGELMETLIGADVULSGARACA 288  
DB 46 LR-----STYAKLTPQOKTQV-----ARHPEKPHLAYISTLIBEDFTPLAGDRAFA 91  
QY 289 LSPAVVALARIIGRPVVLIGQDR-----RFTLG---POLRFARRGISLAREINLP 337  
DB 92 EBEALITIGIGRRFRAASVMIIGHEKHDTETRLKHNFGMAKPEGYKAKAKLMEMADHPQVP 151  
QY 338 IVSIIIDTSGAELSOAAEELGIASSIARTLSKLIDAPLPTVSIIIGQVGGALAMLPADL 397  
DB 152 ITTAVDTGAYPGVDAEARGQABAIARSIETCTNVRVPLVSIIIGGGSGGALATATGNT 211

QY 398 VYAEENMLSLAPREGASAILFRDTNHAETIERQGVQAHALLSQGLIDGIVAE----TE 453  
DB 212 VLMLEHAIYVISPEGCASILMSAENAKAQAQRLTADLKHLSIISDVPEPMOGAH 271  
QY 454 HVEVEILGITSNA-----LSELNPNPERAGDSRFTREER 488  
DB 272 RNPDLMMQTLMSALRDLISGLDGVLRARREKREKLWGR 313

## RESULT 9

US-09-328-352-8167  
Sequence 8167, Application US/09328352  
Patent No. 6562958

GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 8167  
LENGTH: 279  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-8167

Query Match 11.7%; Score 286; DB 4; Length 279;  
Best Local Similarity 31.3%; Pred. No. 6.2e-21;  
Matches 88; Conservative 47; Mismatches 106; Indels 40; Gaps 8;

QY 232 IQPV-EATDRSPITTPGVALPWMEAKRSRDPQPG---IGELMETLIGADVULSGARA 286  
DB 4 IRPKKATQSKAMTTVOIA-----RHPEKPHLAYIGELFTTFDA---LHGDRL 49  
QY 287 GALSPAVVALARIIGRPVVLIGQDR-RFT-----LGPOLRFARRGISLAREIN 335  
DB 50 FGDDGAVWGGLARLDGQPVWVIGQHRGRSTREKLNFGMCNPEGYKKSQRLDMARFN 109  
QY 336 LPIVSIIDTSGAELSOAAEELGIASSIARTLSKLIDAPLPTVSIIIGQVGGALAMLP 395  
DB 110 LPVFTPIDTMGAYPGVDAEARGQAEATISLAQLSLKVPVIAVLIDEGSGGALGIGVA 169  
QY 396 DIVVAENAMLSALPREGASAILFRDTNHAETIERQGVQAHALLSQGLIDGIVAE---- 451  
DB 170 DRVMLSHSYISVSPGCASILMTKAEKAAQSEALGLTADKLSIGIVYVDEGECA 229  
QY 452 ---TEHVEVEILGITSNALSEL---DNNPERAGDSRFTRE 486  
DB 230 HLDPERVWQNLKVVLKQALDELPLMDANRCEARVQRLMKF 270

## RESULT 10

US-09-134-001C-3786  
Sequence 3786, Application US/09134001C  
Patent No. 6380370

GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3786  
LENGTH: 315  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis



## RESULT 13

US-08-611-107-12

Sequence 12, Application US/08611107

Patent No. 5801233

GENERAL INFORMATION:

APPLICANT: Haselkorn, Robert

APPLICANT: Gornicki, Piotr

TITLE OF INVENTION: NUCLEIC ACID COMPOSITIONS ENCODING

TITLE OF INVENTION: ACETYL-COA CARBOXYLASE AND USES

TITLE OF INVENTION: THEREFOR

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Arnold, White &amp; Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: United States of America

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/611,107

FILING DATE: Concurrently Herewith

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US SN 07/956,700

FILING DATE: 02-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US PCT/US93/09340

FILING DATE: 30-SEP-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US SN 08/422,560

FILING DATE: 14-APR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Kitchell, Barbara S.

REGISTRATION NUMBER: 33,928

REFERENCE/DOCKET NUMBER: ARCD:221

TELECOMMUNICATION INFORMATION:

TELEPHONE: (512) 418-3080

TELEFAX: (512) 474-7577

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 327 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-08-611-107-12

Query Match 10.3%; Score 252.5; DB 1; Length 327;  
Best Local Similarity 25.2%; Pred. No. 2.2e-17;  
Matches 86; Conservative 62; Mismatches 138; Indels 55; Gaps 9;

171 LTFAPGAGIGFLGRVVELTGHALPDGVQAEMLVKTGVIDIVSPLOLRAAVAKTLK 230  
12 LEFEKPLVE--LEERTIQIRT-----LAADNOVD--VSGIQOLEKRAIQLR--- 54  
231 VIQPEATDRSPPTPGVALPWEAIAASRDPORPGIGIMETLIGADVVKLSGARAGALS 290  
55 -----REIFSNLSPPAQRIQV-----ARHPRRSTLDYIQAISDEWIELHGDNGSDD 101  
291 PAVVALARIGRPPVVLIGOD-----RRFTL-GPQELRPARRGISLARLNLPIV 339  
102 LALVGVALDQCPVVFILGHQKGRDTKDNVLRNFGWASPGYRKALRLMEHADRFGMPIIL 161  
340 SIIDTSGAELSQAEEELGIASSIARTLSKLIDAPLPTVSVIIGQVGGGALLAMLPAVLVY 399  
162 TFIIDTPGAVGASAEELGGGGAIVANLRMFRFSVPLCTVIGGSGGAGLIGVGRLL 221  
400 AAEANMLSALPPEGASAILFRDTNHAETIERQGVQAHALLSQGLIDGIVA----- 450

Db 222 MFEHSVYTVASPEKCAIILWRDAGKAAQAQAALKITARDLKQGIIDEITTEPLGASHA 281  
QY 451 --ETEHFVEELIGTISNALSIEDNNPERAGDRFTFEEL 489  
Db 282 PLETAQSLRQVL--LRHLKDLQALSPAQLREORYQKFRQL 319

## RESULT 14

US-08-422-560A-12

Sequence 12, Application US/08422560A

Patent No. 5910626

GENERAL INFORMATION:

APPLICANT: Haselkorn, Robert

APPLICANT: Gornicki, Piotr

TITLE OF INVENTION: ACETYL-COA CARBOXYLASE COMPOSITIONS AND

TITLE OF INVENTION: METHODS FOR USE

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Arnold, White &amp; Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: TX

COUNTRY: USA

ZIP: 77210-4433

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/422,560A

FILING DATE: 14-APR-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/956,700

FILING DATE: 02-OCT-1992

ATTORNEY/AGENT INFORMATION:

NAME: Wilson, Mark B.

REGISTRATION NUMBER: 37,259

REFERENCE/DOCKET NUMBER: ARCD:152/WIM

TELECOMMUNICATION INFORMATION:

TELEPHONE: 512-418-3000

TELEFAX: 512-474-7577

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 327 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-08-422-560A-12

Query Match 10.3%; Score 252.5; DB 2; Length 327;  
Best Local Similarity 25.2%; Pred. No. 2.2e-17;  
Matches 86; Conservative 62; Mismatches 138; Indels 55; Gaps 9;

171 LTFAPGAGIGFLGRVVELTGHALPDGVQAEMLVKTGVIDIVSPLOLRAAVAKTLK 230  
12 LEFEKPLVE--LEERTIQIRT-----LAADNOVD--VSGIQOLEKRAIQLR--- 54  
231 VIQPEATDRSPPTPGVALPWEAIAASRDPORPGIGIMETLIGADVVKLSGARAGALS 290  
55 -----REIFSNLSPPAQRIQV-----ARHPRRSTLDYIQAISDEWIELHGDNGSDD 101  
291 PAVVALARIGRPPVVLIGOD-----RRFTL-GPQELRPARRGISLARLNLPIV 339  
102 LALVGVALDQCPVVFILGHQKGRDTKDNVLRNFGWASPGYRKALRLMEHADRFGMPIIL 161  
340 SIIDTSGAELSQAEEELGIASSIARTLSKLIDAPLPTVSVIIGQVGGGALLAMLPAVLVY 399  
162 TFIIDTPGAVGASAEELGGGGAIVANLRMFRFSVPLCTVIGGSGGAGLIGVGRLL 221  
400 AAEANMLSALPPEGASAILFRDTNHAETIERQGVQAHALLSQGLIDGIVA----- 450

Db 222 MEHSVYTVASPEACASILMRDAGKAQAALKTITARDLKQGLIDBITTEPLGAGNSA 281  
QY 451 --ETEHFVEILIGTISNALSELNPNPERAGRDSRFTREEL 489  
Db 282 PLETAQSLRQYL--LRHLKDLQALSPAQLREQRVQKFRQL 319

## RESULT 15

US-08-468-793-12  
Sequence 12, Application US/08468793  
Patent No. 6177267

## GENERAL INFORMATION:

APPLICANT: Haselkorn, Robert  
APPLICANT: Gornicki, Piotr  
TITLE OF INVENTION: ACETYL-CoA CARBOXYLASE COMPOSITIONS AND  
METHODS OF USE  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P. O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,793  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/422,560  
FILING DATE: 14-APR-1995  
APPLICATION NUMBER: US SN 07/956,700  
FILING DATE: 02-OCT-1992  
CLASSIFICATION: 800  
APPLICATION NUMBER: PCT/US93/09340  
FILING DATE: 30-SEP-1993  
CLASSIFICATION: 800

## ATTORNEY/AGENT INFORMATION:

NAME: Kitchell, Barbara S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: ARCD:152/KIT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924

## INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:  
LENGTH: 327 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-468-793-12

## Query Match 10.3%; Score 252.5; DB 3; Length 327;

Best Local Similarity 25.2%; Pred. No. 2.2e-17;  
Matches 86; Conservative 62; Mismatches 138; Indels 55; Gaps 9;

QY 171 LTFAPGAGIIGFLGRVVELTGHALPDGQOENLVKTVGIDGIVSPQLAAVAKTLK 230  
Db 12 LEFEKPLVE--LEERITQIRT-----LADNQVD--VSGQIQULEARAIIQLR--- 54  
QY 231 VIQPEATDRSPPTPGVALPVMELIARSDPQPGIGMETTLGADVVKLSGARAGALS 290  
Db 55 -----REIFSNLSPAQRIQV-----ARHPRRPSTLDYIOAISDEWIELHGDHNGSD 101  
QY 291 PAVVALARIGRPFVLIGD-----RRFTL-GPQELRFARRGISLARELNPV 339

Db 102 LALVGVGALDQGPVVFIGHQGRDPTKONVLENFGMASPGYRKALRLMEHADRFGMPIL 161  
QY 340 SIIDTSGAELSGAAEELGIASSIARTLSKLDAIPPTVSIIIGVGGALAMEPADLVY 399  
Db 162 TPIIDPGAVAGVSABELGGEALVAVNLEMPRFSVPIICTVIGEGSGGALGIGVDRL 221  
QY 400 AENAMLSALPPEGASATLPRDTNHAELIERQVQAHALLSQGLIDGIVA----- 450  
Db 222 MEHSVYTVASPEACASILMRDAGKAQAALKTITARDLKQGLIDBITTEPLGAGNSA 281  
QY 451 --ETEHFVEILIGTISNALSELNPNPERAGRDSRFTREEL 489  
Db 282 PLETAQSLRQYL--LRHLKDLQALSPAQLREQRVQKFRQL 319

Search completed: November 13, 2003, 06:58:22  
Job time : 32 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 06:40:30 ; Search time 366 Seconds  
(without alignments)  
244.909 Million cell updates/sec

Title: US-10-024-370-3

Perfect score: 2448  
Sequence: 1 VKRPPVWVGMEHTSALT.....NNPAGRDSRFRERLAQ 491

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: Published Applications\_Aa.\*  
2: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2448	100.0	491	US-10-024-370-3	Sequence 3, Appl1
2	2445	99.9	491	US-09-738-626-4429	Sequence 4429, Ap
3	1047	42.8	495	US-09-712-363-181	Sequence 181, App
4	487.5	19.9	465	US-10-156-761-13252	Sequence 13252, A
5	400	16.3	288	US-09-815-242-10925	Sequence 10925, A
6	384	15.7	289	US-09-815-242-11385	Sequence 11385, A
7	384	15.7	289	US-09-815-242-11385	Sequence 11385, A
8	375	15.3	330	US-09-815-242-13952	Sequence 13952, Appl
9	374	15.3	304	US-09-815-242-10230	Sequence 10230, A
10	368	15.0	296	US-09-815-242-11212	Sequence 11212, A
11	343.5	14.0	285	US-09-815-242-12618	Sequence 12618, A
12	340.5	13.9	252	US-09-815-242-5424	Sequence 5424, App
13	334	13.6	288	US-09-815-242-13365	Sequence 13365, A
14	334	13.6	288	US-09-815-242-13618	Sequence 13618, A
15	329.5	13.5	290	US-09-815-242-11898	Sequence 11898, A

16	291	11.9	532	15	US-10-156-761-10868	Sequence 10868, A
17	288	11.8	315	9	US-09-815-242-11014	Sequence 11014, A
18	286	11.7	531	15	US-10-156-761-10240	Sequence 10240, A
19	285	11.6	527	12	US-10-045-612A-31	Sequence 31, Appl
20	277	11.3	309	9	US-09-815-242-5892	Sequence 5892, Ap
21	275	11.2	314	9	US-09-815-242-12908	Sequence 12908, A
22	275	11.2	526	15	US-10-187-267A-47	Sequence 47, Appl
23	270	11.0	266	9	US-09-815-242-4964	Sequence 4964, App
24	269	11.0	319	9	US-09-815-242-13897	Sequence 13897, A
25	268.5	11.0	255	9	US-09-815-242-13366	Sequence 13366, A
26	268.5	11.0	255	9	US-09-815-242-13619	Sequence 13619, A
27	268.5	11.0	514	15	US-10-156-761-9840	Sequence 9840, Ap
28	266	10.9	319	9	US-09-815-242-10048	Sequence 10048, A
29	265.5	10.8	324	9	US-09-841-132-496	Sequence 496, App
30	265	10.8	262	9	US-09-815-242-10926	Sequence 10926, A
31	261.5	10.7	484	9	US-09-925-301-1140	Sequence 1140, App
32	261.5	10.7	484	15	US-10-106-698-5893	Sequence 5893, Ap
33	257	10.5	312	9	US-09-815-242-11339	Sequence 11339, A
34	257	10.5	312	9	US-09-815-242-11511	Sequence 11511, A
35	254	10.4	516	10	US-09-738-626-6665	Sequence 6665, Ap
36	252.5	10.3	324	9	US-09-841-132-566	Sequence 566, App
37	252.5	10.3	327	9	US-09-767-479-12	Sequence 12, Appl
38	249.5	10.2	316	9	US-09-815-242-11934	Sequence 11934, A
39	241	9.8	537	10	US-09-738-626-4284	Sequence 4284, A
40	218.5	8.9	532	15	US-10-156-761-11401	Sequence 11401, A
41	217.5	8.9	593	15	US-10-224-539A-12	Sequence 12, Appl
42	215.5	8.8	543	10	US-09-738-626-4285	Sequence 4285, Ap
43	207.5	8.5	543	15	US-10-259-648-2	Sequence 2, Appl
44	190.5	7.8	563	15	US-10-224-539A-10	Sequence 10, Appl
45	190.5	7.8	563	15	US-10-205-823-254	Sequence 254, App

## ALIGNMENTS

RESULT 1  
US-10-024-370-3  
; Sequence 3, Application US/10024370  
; Application No. US20020142405A1  
; GENERAL INFORMATION:  
; APPLICANT: TILG, YOONNE  
; APPLICANT: ELKMANN, BERND  
; APPLICANT: EGGELING, LOTMAR  
; APPLICANT: SAHM, HERMANN  
; APPLICANT: MCKEL, BETTINA  
; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF L-AMINO ACIDS BY  
; TITLE OF INVENTION: PERMENTATION AND NUCLEOTIDE SEQUENCES CODING FOR THE  
; FILE REFERENCE: 21123-284139-MAS  
; CURRENT APPLICATION NUMBER: US/10/024,370  
; PRIOR FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 09/362,899  
; PRIOR FILING DATE: 1999-07-29  
; PRIOR APPLICATION NUMBER: DE 199 24 365.4  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 3  
; LENGTH: 491  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-10-024-370-3  
Query Match 100.0%; Score 2448; DB 14; Length 491;  
Best Local Similarity 100.0%; Pred. No. 1.1e-217;  
Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VKRPPVWVGMEHTSALTLDVDPDSFISNMTPEYDUNQGYATIRASAKCD 60  
DB 1 VKRPPVWVGMEHTSALTLDVDPDSFISNMTPEYDUNQGYATIRASAKCD 60  
QY 61 ESVTGEVTEGIPAVVILSDPSFGSLGVASVRIKAIHRAETELPLVPSASGCA 120

Db 61 ESVITGEGTVEGIPVAVILSDSFSLGSLGTVASVRIMKAIHRATELKLPLVSPASGA 120  
QY 121 RMOEDNRAFWVMVSTTAAVQHRRAHLPLVYLNPMTGAMASWSSGHLTFAPGAOI 180  
Db 121 RMOEDNRAFWVMVSTTAAVQHRRAHLPLVYLNPMTGAMASWSSGHLTFAPGAOI 180  
QY 181 GFLGRVVELTTGHALPDGVQQAENLVKTGVIDGIVSPLOLRAAVATLKVIOVEATDR 240  
Db 181 GFLGRVVELTTGHALPDGVQQAENLVKTGVIDGIVSPLOLRAAVATLKVIOVEATDR 240  
QY 241 FSPPTPGVALPVMEAIARSRDPOPGIGEMETLGADVVKLSGARAGALSPAVVALARI 300  
Db 241 FSPPTPGVALPVMEAIARSRDPOPGIGEMETLGADVVKLSGARAGALSPAVVALARI 300  
QY 301 GGRPVVLIGODRRFTLGPOLRFARRGISLARELNLPVSIIDTSGAELSQAAEELGAS 360  
Db 301 GGRPVVLIGODRRFTLGPOLRFARRGISLARELNLPVSIIDTSGAELSQAAEELGAS 360  
QY 361 STARTLSKLIDAPLPTVSVIIIGGVGGALAMPADLVYAENAMLSALPEGASALIFR 420  
Db 361 STARTLSKLIDAPLPTVSVIIIGGVGGALAMPADLVYAENAMLSALPEGASALIFR 420  
QY 421 DTNHAETIERQGVQAHALLSQGLIDGIVATEHFEVEIIGTISNALSELNNPERAGR 480  
Db 421 DTNHAETIERQGVQAHALLSQGLIDGIVATEHFEVEIIGTISNALSELNNPERAGR 480  
QY 481 SRTFRFERLAQ 491  
Db 481 SRTFRFERLAQ 491

## RESULT 2

US-09-738-626-4429  
; Sequence 4429, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4429  
; LENGTH: 491  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4429

Query Match 99.9%; Score 2445; DB 10; Length 491;  
Best Local Similarity 99.8%; Pred. No. 2e-217;  
Matches 490; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VEKRPPTVMGMEHTSALTLLIDSVDDPSFISWNETPOYDNNNGYATLEBARAKACD 60  
Db 1 MEKRPPTVMGMEHTSALTLLIDSVDDPSFISWNETPOYDNNNGYATLEBARAKACD 60  
QY 61 ESVITGEGTVEGIPVAVILSDSFSLGSLGTVASVRIMKAIHRATELKLPLVSPASGA 120

Db 61 ESVITGEGTVEGIPVAVILSDSFSLGSLGTVASVRIMKAIHRATELKLPLVSPASGA 120  
QY 121 RMOEDNRAFWVMVSTTAAVQHRRAHLPLVYLNPMTGAMASWSSGHLTFAPGAOI 180  
Db 121 RMOEDNRAFWVMVSTTAAVQHRRAHLPLVYLNPMTGAMASWSSGHLTFAPGAOI 180  
QY 181 GFLGRVVELTTGHALPDGVQQAENLVKTGVIDGIVSPLOLRAAVATLKVIOVEATDR 240  
Db 181 GFLGRVVELTTGHALPDGVQQAENLVKTGVIDGIVSPLOLRAAVATLKVIOVEATDR 240  
QY 241 FSPPTPGVALPVMEAIARSRDPOPGIGEMETLGADVVKLSGARAGALSPAVVALARI 300  
Db 241 FSPPTPGVALPVMEAIARSRDPOPGIGEMETLGADVVKLSGARAGALSPAVVALARI 300  
QY 301 GGRPVVLIGODRRFTLGPOLRFARRGISLARELNLPVSIIDTSGAELSQAAEELGAS 360  
Db 301 GGRPVVLIGODRRFTLGPOLRFARRGISLARELNLPVSIIDTSGAELSQAAEELGAS 360  
QY 361 STARTLSKLIDAPLPTVSVIIIGGVGGALAMPADLVYAENAMLSALPEGASALIFR 420  
Db 361 STARTLSKLIDAPLPTVSVIIIGGVGGALAMPADLVYAENAMLSALPEGASALIFR 420  
QY 421 DTNHAETIERQGVQAHALLSQGLIDGIVATEHFEVEIIGTISNALSELNNPERAGR 480  
Db 421 DTNHAETIERQGVQAHALLSQGLIDGIVATEHFEVEIIGTISNALSELNNPERAGR 480  
QY 481 SRTFRFERLAQ 491  
Db 481 SRTFRFERLAQ 491

## RESULT 3

US-09-712-363-181  
; Sequence 181, Application US/09712363  
; Patent No. US20020164588A1  
; GENERAL INFORMATION:  
; APPLICANT: Eisenberg, David  
; APPLICANT: Rotstein, Sergio H.  
; APPLICANT: Marcotte, Edward M.  
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
; FILE REFERENCE: 07419-032001  
; CURRENT APPLICATION NUMBER: US/09/712, 363  
; CURRENT FILING DATE: 2000-11-13  
; PRIOR APPLICATION NUMBER: PCT/US00/02246  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: 60/179, 531  
; PRIOR FILING DATE: 2000-02-01  
; PRIOR APPLICATION NUMBER: 60/117, 844  
; PRIOR FILING DATE: 1999-01-29  
; PRIOR APPLICATION NUMBER: 60/118, 206  
; PRIOR FILING DATE: 1999-02-01  
; PRIOR APPLICATION NUMBER: 60/126, 593  
; PRIOR FILING DATE: 1999-03-26  
; PRIOR APPLICATION NUMBER: 60/134, 093  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/134, 092  
; PRIOR FILING DATE: 1999-05-14  
; PRIOR APPLICATION NUMBER: 60/165, 124  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/165, 086  
; PRIOR FILING DATE: 1999-11-12  
; NUMBER OF SEQ ID NOS: 292  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 181  
; LENGTH: 495  
; TYPE: PRT  
; ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-181

Query Match 42.8%; Score 1047; DB 10; Length 495;  
Best Local Similarity 47.8%; Pred. No. 4.5e-88;

	Matches	233;	Conservative	73;	Mismatches	153;	Indels	28;	Gaps	8;
Oy	23	SULDDDSFISMNETQYDNLNNGYAETLERARSKKCBESVYTGGTVEGIPAVITLSE	:	:	:	:	:	:	:	82
Dd	12	AALDGSGFVSWSDEFLAAPPVADSVARELAATAATGADESVOGTGCGRFGGRVAAVACEF	:	:	:	:	:	:	:	71
Oy	83	SPLGSGLGVASVRIMKAIIHRAATEKLPLTVSPASGARGMOEDNAFMMWISITAAYGRH	:	:	:	:	:	:	:	142
Dd	72	DPLGGSIGVAHAERITTAVERITAERLPULMSPSGGSTRMOEGIVAFLOMKIAAALQLH	:	:	:	:	:	:	:	131
Oy	143	REAHLPFLVYLRLNPMTMGAMASWGSSGHLLTFEAPGOIGFLGRPVVELTTGHALPDGYQC	:	:	:	:	:	:	:	202
Dd	132	NQARLPFLVYLRLHPTTGCVFASWGSLSGHULTVAEPGALLGFLGRPYELLYXDPPFSGQTI	:	:	:	:	:	:	:	191
Oy	203	AENLVKTGVINGISVPFLQRAAVAKTLKYI---OPVATDRFSFTPGVA--LPWEAI	:	:	:	:	:	:	:	256
Dd	192	AENLRHRHIIIGVALIDLRLRPMLDRALTVLIDAPPFLPA-----FOTPAPVDVPTMWSV	:	:	:	:	:	:	:	246
Oy	257	ARSRDPOGPGIGEIMETLGADVVKLSGARAGALSFAVVALARIGGRVVYLIGDRRRP-	:	:	:	:	:	:	:	314
Dd	247	VASRRPDPFGVRQILRHGATBRVLISGDGEAAATTL-LALARRGGQTIVVLVGQORAVGC	:	:	:	:	:	:	:	305
Oy	315	--TLGQELRFARRGISLARBLMLPIYSIIDTSAGELSQAABELGISIARTLSKLIE	:	:	:	:	:	:	:	371
Dd	306	GGSTGFPALAEARRGMALELCPELVLTIDAAPALSAABOGGLAQOIAHCIAELVT	:	:	:	:	:	:	:	365
Oy	372	APLPVSVYITIGCGGGCLAMLPADLVYAAENAMLSALPRCSAISILFRDTNHAAEITER	:	:	:	:	:	:	:	431
Dd	366	LDTPLVSLILIQSGSGPALAMLPRDVYLAALHGWLAPLPPEGASAIYFRDTNHAAEILAA	:	:	:	:	:	:	:	425
Oy	432	QGVOAHALLSQGLDIGIVAETHFVEEITLG---TISNMLSELDN-----NPERAGDSR	:	:	:	:	:	:	:	482
Dd	426	QGHSADLLKSIGVIDTYIPEIFDADEPIEFALRLSNIAALEVHLKRKIAPERILA-TR	:	:	:	:	:	:	:	483
Oy	483	FTRFERI 489	:	:	:	:	:	:	:	
Dd	484	LQRYYRI 490	:	:	:	:	:	:	:	
 RESULT 4 US-10-156-761-13252 Sequence 13252, Application US/10156761 Publication No. US20030119018A1 GENERAL INFORMATION: APPLICANT: OMURA, SATOSHI APPLICANT: IKEDA, HARUO APPLICANT: ISHIKAWA, JUN APPLICANT: HORIKAWA, HIROSHI APPLICANT: SHIBA, TADAYOSHI APPLICANT: SAKAKI, YOSHIYUKI APPLICANT: HATTORI, MASAHIRA TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES FILE REFERENCE: 249-262 CURRENT APPLICATION NUMBER: US/10/156, 761 PRIOR FILING DATE: 2002-05-29 PRIOR APPLICATION NUMBER: JP 2001-204089 PRIOR FILING DATE: 2001-05-30 PRIOR APPLICATION NUMBER: JP 2001-272697 PRIOR FILING DATE: 2001-08-02 NUMBER OF SEQ ID NOS: 15109 SEQ ID NO 13252 LENGTH: 465 TYPE: PR1 ORGANISM: Streptomyces avermitilis US-10-156-761-13252										

```

0Y      78      ILSPFSFGGSLGTVAASRIMKAIHRAIEIKLPLVSPSGARMOEDNRAFFVMWVSITA 137
       71      IAFEGFGGSLGERTGRLEAAITYAARHLRPVSVATGSSRMOEGMLALTQOLORVAR 130
0Y      138      AVQRREARHLPVLYLRNPTMGAMASWGSSGHLTFAEBGAOIGFLGPRVVELTTGHALP 197
       131      QSALTFRAGLPLQALVLRDPTTGGGMATLIGAGDVVLTLPGVGVFGASRV-----RPP 183
0Y      198      D---GVQQAENLVKTVGVIDGIVSPQLFAAAVAKTKLVIOPVZATDRFSSTTTCVALPWAE 254
       184      DADPAAYTAOQLAAGSIDAVVRPEOLEKRETVGLMTALL---TABSTEEPAAPPALGTS 239
0Y      255      -----AIKSRNDPORGGIEMETIGADVVKLSGARAGALSAPVRVAL----- 297
       240      RLPRDTGMNAVGRASPERPAQAYIDAYFTTRRAIAGDRCGSTDGMLCGFGFSSGSSG 299
0Y      298      --ARIGRPVVLIGDRRFTLPGOLEFARRGISLABRLNPIVSIIDTSGNELSOAAE 354
       300      SGSEHGGRVAVVAAQTSTAT-RPAGYKTAATRLRLADRLGIPVLTIVTTPGANDAABAE 358
0Y      355      ELGIASSIARTLSKLIIDAPLPTVSVITIGCVGGGGLAMLPADLVVAAENAMISALPPEGA 414
       359      RQGGAALAEIFGAVATARTPLTLLIGEGSGGGLATAAPGNTATPDSYSVIAPBEIA 418
0Y      415      SAILFRDTHAAEIIERGQVOAHMLISGGLIDGIV 449
       419      AAILKRPPEELPAGTADQLRVRPQDLVALGAVRGIV 453

```

```

RESULT 5
US-09-815-242-10925
: Sequence 10925, Application US/09815242
: Patent No. US2002061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zykkind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITTA.01A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/131,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 10925
: LENGTH: 288
: TYPE: PRF
: ORGANISM: Enterococcus faecalis
: US-09-815-242-10925

```

```

Query Match      19.9%;   Score 487.5;   DB 15;   Length 465;
Best Local Similarity 29.9%;   Pred. No. 2,46; 36;
Matches 136;   Conservative 66;   Mismatches 208;   Indels 45;   Gaps 7;

QY      18  LTLIDSVLDPSFISKNETPQYDNLNCGYAELEAKAKKCDSEVITGEGTVEGIPAV 77
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db       21  LPVAGAAVAPGQPSW-----QGIDSSRAATATETGDEBSSVCGGRRATIDTPAVL 70

```

Query Match	16.3%	Score 400;	DB 9;	length 288;
Best Local Similarity	42.6%	Pred. No. 1.4e-28;		
Matches	89;	Conservative	39;	Mismatches 79;
			Indels	2;
			Gaps	2

[illegible]

```

RESULT 6
US-09-815-242-11385
Sequence 11385, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITR.011A
CURRENT APPLICATION NUMBER: US/09/815.242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/109,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11385
LENGTH: 289
TYPE: PRT
ORGANISM: Helicobacter pylori
US-09-815-242-11385

```

	Query Match	15.7%	Score 384	DB 9	Length 289	
	Best Local Similarity	39.3%	Pred. No. 4.3e-27			
	Matches	84	Conservative	40	Mismatches	72
					Indels	18
					Gaps	2
QY	26	DPDSEISWNETPQYDNLNOGVAFETTERARSKKCDSEVITGEGTGVIPAVILDSFSL	85			
		: :	: : :	: :	: : :	
Db	85	DPLNFVD-----KESYKQRIKTKYEKINRRSSVTSIGAKINRMPLQIVDFDSFM	134			
QY	86	GGSGTSAVSRIMKIIHRATELKLPLVSPASGAGARMOEDNFAFVMQNSITAAVQHRREA	145			
		: :	: : :	: :	: : :	
Db	135	GGSGTSGVEGEKIVRAINRAVAREALLIVASAGSGARMOESTSLMQAKTSAALNRLSEA	194			
QY	146	HLPLVLVLRNPMGMAMASWGSSGHLTFAPRQAQGLFGRVVELLTGHALDDGVQAQEN	205			
		: :	: : :	: :	: : :	
Db	195	KLPLTSLSDPFGVGVASAFALGDLIIAEPAMTGFAPAPRIKQTIGADIDEGGQVTFE	254			

```
Qy      206 LAKTGVADIGVBPLOLRRAVA-----KTLKV 231
          ||| | : : |||
Db      255 LLEHGLDNIYHRKDCLKTSLDLAMTHYSKI 288
```

```

RESULT 7
US-09-895-913A-98
; Sequence 98, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleantchous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 98
; LENGTH: 289
; TYPE: prt
; ORGANISM: Helicobacter pylori
US-09-895-913A-98

```

Query Match	15.7%	Score 384	DB 10	Length 289
Best Local Similarity	39.3%	Pred. No. 4.3e-27		
Matches	84	Conservative 40	Mismatches 72	Indels 18
			Gaps	2
QY	26	DPDSFISMETPQYDNLNQGYYAETLERARSAKCDSESYITBEGYEGIPVAVILSDSFSL	85	
DB	85	DPLNFVD-----KESYKORIKKKYEKTNPPSYISEAKINRMLQIVPDFSFM	134	
QY	86	GGSLGTVA SVIRKAIHRA TELKCIPLVSPASGARGMEDNRA FVMWYSITAAVORHBEA	145	
DB	135	GGSLGASVEGEKTVAINRAVAKREALLVLSASGARMOESTYSLMQAKTSALNRLESEA	194	
QY	146	HLPLVYLRLNDPTMGAMASWSSGHLTFPAEFGAIGFTGPRVVELTTGHALPDGVOQAE N	205	
DB	195	KLPLFISLSDPTYGVSASFALFDLLIAEBGAMIGFAGPRVVIKQTIGADIPGEGOTAEF	254	
QY	206	LYTKGVIDGIVSPQLRLRAVA-----KTLKV	231	
DB	255	LLRHGLIDMTIVHRDKLKKTLSDLIAMMTHKTSKI	288	

```

RESULT 8
US-09-815-242-13952
: Sequence 13952, Application US/09815242
: Patent No. US20020061559A1
: GENERAL INFORMATION:
: APPLICANT: Hasselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELTRA, 011A
: CURRENT FILING DATE: US/09/815,242
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21

```

```

; PRIOR APPLICATION NUMBER: 60/206, 848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207, 727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242, 578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253, 625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257, 931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269, 308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13952
; LENGTH: 330
; TYPE: prt
; ORGANISM: Salmonella typhi
US-09-815-242-13952

```

Query Match	15.3%	Score 375;	DB 9;	Length 330;
Best Local Similarity	36.0%;	Pred. No. 3.6e-26;		
Matches	85;	Conservative	49;	Mismatches 92;
			Indels	10;
			Gaps	3

Qy 17 ALTLISVLDPDPSFIEMNEPQY-DLUNGVAETLERASRKAKODESITBEGVEGIPV 75  
Db 95 SLVBLSSELEPPKVLFRDSKKYDRLASAQETGK-----DALVNGKLTTHGPP 146  
Qy 76 AVILSDFSFLGSLGTVASVIRIKALHRAETELKPLVSPASGARGMOEDRAFAVMVSI 135  
Db 147 VAAPAFEPAGFGMSGVSGARFVRAVEQLLEJDNCPVCSGARGMOEDALMSLMQAKT 206  
Qy 136 TAAVQHRREHLPLVLYLRNPTVGGAMASWSSGHLTPAEPAQIGTGLPRVVELTGH 195  
Db 207 SAALAKQOEGLFYISVLTDPMTGGVSASFAMGLDNLIAEPALILGFPGPRVIEQVREK 266  
Qy 196 LPPGVQOAEULVYTGVIDGIVSPLODRRAVAKTLKVIQVEATDRFSPTTGVALP 251  
Db 267 LPFGFQSEFLIEKGLIMIVRRPEERLKLASILKLKNNLPDPNDAP-REGVAVP 321

```

RESULT 9
US-09-815-242-10230
; Sequence 10230, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlssen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; FILE REFERENCE: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21, 078
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

```

```

; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10230
; LENGTH: 304
; TYPE: prt
; ORGANISM: Escherichia coli
; OS-09-815-242-10230

```

Query Match	15.3%;	Score 374;	DB 9;	Length 304;
Best Local Similarity	36.0%;	Pred. No. 3.9e-26;		
Matches	85;	Conservative	49;	Mismatches 92;
			Indels	10;
			Gaps	3

QY 17 ALTLIISVDDPDSFISMNETPOL -ONLNGVAETLERARSXAKDEBSYITGEVGEIPV 75  
69 SLVELDSSELEPKDVLKFRDSKKT KDRLASAQETGEK-----DALVWNGKGLYQHPV 120

QY 76 AVILSDPSFLGSLGTVASVRIMKA IHRATEIKUPLVSPASGARMQEDNRFAVMVYSI 139  
121 VAAAFEFAPMGSMGVSVCARFVRAVEQLEDNCPICFSASGARMQEDALMSLMQAKT 180

QY 136 TAAVOCHREANHLPELVLEARNPTMGAMMSKSSGHLTAEFGAQIGFLGRPVAVELTTGHA 195  
181 SAALAMQORCGPYISVLTDPMTMGVSSAFAMEGLNIAEBKALIGAGRPVIEQTVREK 240

QY 196 LEPGVOQAENLWKTVGIVIGIVSPLOLRAAVAKTLKVQPVBATDRSPPTPGVALP 251  
241 LEPGFRSSEFLIEKADIDVIRAPRPBKTLASILAKMNLPAAPNPEAP-REGVAVP 295

RESULT 10  
US-09-815-242-11212

; Sequence 11212, Application US/09815242  
; Patent No. US20020061569A1

```

1  APPLICANT: Haeselbeck, Robert
2  APPLICANT: Ohlсен, Kari L.
3  APPLICANT: Zyskind, Judith W.
4  APPLICANT: Wall, Daniel
5  APPLICANT: Tralock, John D.
6  APPLICANT: Carr, Grant J.
7  APPLICANT: Yamamoto, Robert T.
8  APPLICANT: Xu, H. Howard
9  TITLE OF INVENTION: Identification of Essential Genes in
10 TITLE OF INVENTION: Procharotes
11 FILE REFERENCE: ELITRA.01A
12 CURRENT APPLICATION NUMBER: US/09/815,242
13 CURRENT FILING DATE: 2001-03-21
14 PRIOR APPLICATION NUMBER: 60/131,078
15 PRIOR FILING DATE: 2000-03-21
16 PRIOR APPLICATION NUMBER: 60/206,848
17 PRIOR FILING DATE: 2000-05-23
18 PRIOR APPLICATION NUMBER: 60/207,727
19 PRIOR FILING DATE: 2000-05-26
20 PRIOR APPLICATION NUMBER: 60/242,578
21 PRIOR FILING DATE: 2000-10-23
22 PRIOR APPLICATION NUMBER: 60/253,625
23 PRIOR FILING DATE: 2000-11-27
24 PRIOR APPLICATION NUMBER: 60/257,931
25 PRIOR FILING DATE: 2000-12-22
26 PRIOR APPLICATION NUMBER: 60/269,308
27 PRIOR FILING DATE: 2001-02-16
28 NUMBER OF SEQ ID NOS: 14110
29 SOFTWARE: FastSeq for Windows Version 4.0
30 SEQ ID NO 11212
31 LENGTH: 296
32 TYPE: PRNT
33 ORGANISM: Haemophilus influenzae
34 OS-09-815-242-11212

```

Query Match	15.0%;	Score 368;	DB 9;	Length 296;
Best Local Similarity	36.8%;	Pred. No. 1.3e-25;		
Matches	81;	Conservative	45;	Mismatches 80;
				Indels 14;
				Gaps 3;

Qy	25	LDPSFSISMNEPQY-DNINOGYAETLEBARSHAKODESVITGCEGIVEGIPVAVILSDPS	83
Db	79	LEPKOILKFKDKKKYKDRINNAOKETGEK-----DALITMTGTLXNNPVIIVASNFPA	130
Qy	84	FLGSGLGVAASVRIKAIHRATELKLPLVPSPASGARAQEDNRAPVMMVSIPTAAVORHR	143
Db	131	FMGSGMSGVVGAKFYKAAEKAMEMNCPPVCFSSASGARQEOALLFSLMOAKTSAYVLAQNR	190
Qy	144	EAHLPELVLYLRNPITWGAMASWSSGSHLTFAEBQAOIGFLGPRVELITGHALPDGVQQA	203
Db	191	EKGVPFISVLTDPITLGVSASFAMJGDLNIAEBKALIGFVGPRVIRIQVAREKLPBGFQRS	250
Qy	204	ENLVKTVIYIDGIVSPQLRPAAVAKTLKVQPVKATRRPSP	243
Db	251	EPLEKGAIDMTVKRSEMRQTASVLS-----KLITQPSR	285

```

RESULT 11
US-09-815-242-12618
: Sequence 12618, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 12618
: LENGTH: 285
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
US-09-815-242-12618

```

[illegible]

Qy 188 VELTGHAEDEGVQAENLVKTVGDIVISPLQLRAAVAKTLKVIOV 235  
::|||:::||::||  
Db 236 IEQTINEKEPDDFQTAEFLFHGGQLDKVHRNDMRQTISEILKIHQEV 283

```

RESULT 12
US-09-815-242--5424
; Sequence 5424, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/265,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5424
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5424

```

```

Query Match      13.9%; Score 340.5; DB 9; Length 25;
Match Similarity 35.8%; Pred. No. 3.7e-23;
Base Local      81; Conservative 47; Mismatches 93; Indels 5; Gaps 3;

QY      13  EHTSALT--IDSVLDPDSFISNNETPOYDN-LN-QGYAETTERARSKAKCDESVTIGE 67
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      27  DHIALTPYKKEIEISDEGFTFEDKGMTSANPLDFPSYIEKIKDOOKTKKAAVTGT 86
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      68  GTVEGIPAVAILSDPSFPGSLGTVASVRIKAIHRAITELPLTVSPASGARGMOEDNR 127
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      87  AQLDGMKRGVAMVMSRFRMGSMGVSIEKICRIITDYCTENRLLPILTSASGARGMOEGII 146
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      128  AFVMMVSTIAVQRRREAHLPFLVLYLRNPITMGAMASWGSGHLTFAPGAQIGFLGRV 187
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      147  SLMGQKTSVSLKRRHSDAGLLIYLRHPTTGGVASAFASVGDINLSBPKALIGFAGRVR 206
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      188  VELTTGHALPDGVQDAENLVKTVGIIDGIVSPLOLRPAVAKTLKVIQ 233
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      207  IEQTINERLPDPQTAEFLLEHGQLDKVRHNRDMRQTLSEILKIHQ 252
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-09-815-242-13365
; Sequence 13365, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.

```

```
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 13365
LENGTH: 288
TYPE: PR1
ORGANISM: Streptococcus pneumoniae
US-09-815-242-13365
```

```
Query Match      13.6%; Score 334; DB 9; Length 288;
Best Local Similarity 39.0%; Pred. No. 1.8e-22;
Matches 73; Conservative 36; Mismatches 78; Indels 0; Gaps 0;
```

```
QY 45 GAAETLEBARSKAKODESVITGEVGEIPVAVILSDPSFLGSGIGTVAASVRIMKAHRA 104
DB 96 GYOKKLAMRKERTGHEHENVITGALIKQTVAGIMDSNFTWASNGTVGAKITRLFEYA 155
QY 105 TELKPLIVSPASGARMOEDNRAFMVNSITAAVQRREAHLPVLVLRNPTMGAMAS 164
DB 156 TVEKLPVLFASGARMOEGIMSLMOMAKISAAYKHSNGLPVLITLTPPTGGVTAS 215
QY 165 WSSGHLTFAPPAQOIGFLGRVVELTTGHALPDGVQOENLVKTVGIDGIVSPLOQRAA 224
DB 216 FAMEDDILAEPPQSLVGPAGRVIENTVRESLPEDFOAEFLBHGFDVAIVKRDLPDT 275
QY 225 VAKTLKV 231
DB 276 IASLVR 282
```

```
RESULT 14
US-09-815-242-13618
Sequence 13618, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
```

```
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 13618
LENGTH: 288
TYPE: PR1
ORGANISM: Streptococcus pneumoniae
US-09-815-242-13618
```

```
Query Match      13.6%; Score 334; DB 9; Length 288;
Best Local Similarity 39.0%; Pred. No. 1.8e-22;
Matches 73; Conservative 36; Mismatches 78; Indels 0; Gaps 0;
```

```
QY 45 GAAETLEBARSKAKODESVITGEVGEIPVAVILSDPSFLGSGIGTVAASVRIMKAHRA 104
DB 96 GYOKKLAMRKERTGHEHENVITGALIKQTVAGIMDSNFTWASNGTVGAKITRLFEYA 155
QY 105 TELKPLIVSPASGARMOEDNRAFMVNSITAAVQRREAHLPVLVLRNPTMGAMAS 164
DB 156 TVEKLPVLFASGARMOEGIMSLMOMAKISAAYKHSNGLPVLITLTPPTGGVTAS 215
QY 165 WSSGHLTFAPPAQOIGFLGRVVELTTGHALPDGVQOENLVKTVGIDGIVSPLOQRAA 224
DB 216 FAMEDDILAEPPQSLVGPAGRVIENTVRESLPEDFOAEFLBHGFDVAIVKRDLPDT 275
QY 225 VAKTLKV 231
DB 276 IASLVR 282
```

```
RESULT 15
US-09-815-242-11898
Sequence 11898, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
```



